KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

Quantum mechanical transport towards the optimization of heterostructure tunnel field-effect transistors

Devin Verreck

Supervisors:
Prof. dr. ir. Guido Groeseneken
Prof. dr. ir. Bart Sorée

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering Science (PhD): Electrical Engineering

January 2017

Quantum mechanical transport towards the optimization of heterostructure tunnel field-effect transistors

Devin VERRECK

Examination committee:
Prof. dr. ir. Hugo Hens, chair
Prof. dr. ir. Guido Groeseneken, supervisor
Prof. dr. ir. Bart Sorée, supervisor
Prof. dr. ir. Marc Heyns
Prof. dr. ir. Marian Verhelst
Prof. dr. Michel Houssa

Prof. dr. Mykhailo Povolotskyi
(Purdue University)
Dr. ir. Anne Verhulst
(imec)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor in Engineering Science (PhD): Electrical Engineering

January 2017

Terras licet, inquit, et undas
obstruat; et caelum certe patet,
ibimus illac!

Daedalus
Ovid, *Metamorphoses* VIII

Preface

Long gone are the days when a PhD was the solitary effort of a student locked up in an ivory tower. While the effort has remained, the tower is now mainly glass and concrete, and the student is supported by a whole team of people. I want to take a moment here to thank those people, at imec and elsewhere, who have all had a major role in the completion of this PhD work.

I owe a large dept of gratitude to my daily supervisor, dr. Anne Verhulst. Anne is one of the most rigorous, critical and capable scientists I know. Throughout both my Master's thesis and PhD, she has guided me closely, advising me on research choices, suggesting numerous improvements to my manuscripts and always pushing me to truly understand my simulation results. My admiration for her is great, and she will remain an example for me in my future career.

I also want to thank my co-promotor, prof. Bart Sorée, who regularly helped me with the more quantum mechanical aspects of my modeling. He was always willing to dive into the mathematical weeds and check my derivations. Oftentimes, this was during our Friday PhD theory meetings, where my engineering mind learned to enjoy the particularities of condensed matter physics. At said meetings, prof. Wim Magnus' inquisitive remarks often led me to rethink the motivation behind my approximations.

I am grateful to my promotor, prof. Guido Groeseneken, for providing the topic of both my Master's and PhD thesis and for advocating nanotechnology education at the KU Leuven. He also encouraged me to take on the role of ombudsperson of the nano program, which enhanced my compromise building skills and provided me with an inside look into the workings of the KU Leuven. I want to thank also prof. Marc Heyns for our cooperation in the local nano program, and for spreading his scientific enthusiasm to a new generation of researchers. Thank you Jie and Dimitrios for being my colleagues as ombuds.

There are a lot of people at imec who contributed to this work, both directly and indirectly. In particular, I want to thank Maarten, whose Master's thesis

formed the foundation of this PhD work and whose deep physical insight I could often rely on to help me untangle **k**·**p**-theory. Quentin was my neighbor and experimental counterpart in the TFET team, who diligently worked at connecting our predictions to actual devices. Together with Maarten, Kristof and, later on, Olivier, we had invigorating lunch discussions on the nature of reality, the merits of eugenics and the identity of "de Mol".

Of course all of the current and former members of the TFET team and management also had a large impact: Alireza, Anda, Aaron, Anne VD, Ashish, Frank, Jasper, Mazhar, Nadine, Rita, Salim, William, Quentin and many others. When necessity demanded that I venture into the realm of semiclassical simulations, I could luckily rely on the experience of my TCAD team members Arturo, An, Philippe and Geert. I am also grateful to Claire, Betty and Myriam for arranging all kinds of practicalities, from room bookings to flight schedules. I also thank Geoffrey, for staying helpful and professional when my simulations strained the computational infrastructure beyond its prescribed limits.

Outside of imec, I also spent three months on the rural plains of Indiana at Purdue University, where I briefly joined the group of prof. Gerhard Klimeck and was able to get first-hand experience with NEMO5. I want to specifically thank prof. Michael Povolotskyi, who was very welcoming and who made me realize the importance and non-trivial nature of model calibration. Thank you James, Kuang, Pengyu, Jun, Hesam, David, Jim and all the other group members for supporting me. I particularly enjoyed the soccer sessions, which confirmed my prejudices about Asian and North American soccer skills. This research stay was made possible by the Research Foundation Flanders (FWO).

For providing the funding for this PhD work, I acknowledge the former Agentschap voor Innovatie door Wetenschap en Techniek Vlaanderen (IWT-Vlaanderen), currently the Agentschap Innoveren en Ondernemen. I also thank my examination committee for critically reading the PhD manuscript and suggesting improvements.

Despite the recent expansion rate, there is also still a world outside of imec. I thank my friends from Salco, from Judoclub Bierbeek and from the Leuvense Reddingsclub for the support and diversion they abundantly supplied. I also thank my grandparents for picking me up from school and being there for me every step of the way. Of course, none of this would be possible without my parents. I am thankful for all the opportunities they gave me. This work is dedicated to them.

Abstract

The scaling of the metal-oxide-semiconductor field-effect transistor (MOSFET) has been the driving force for the enormous increase in computational power in everyday digital electronics since the 1960's. Today, this trend is reaching its limits, as the MOSFET supply voltage can no longer be scaled at the same pace as the device dimensions. This is due to a lower limit on the subthreshold swing (SS). As a result, the power density of integrated circuits rises with each new generation, which is eventually untenable.

The tunnel field-effect transistor (TFET) has been developed to break this detrimental evolution. Its operating principle based on band-to-band tunneling (BTBT) enables a low SS and hence low supply voltage operation. Silicon implementations, however, have shown insufficient ON-currents. Research is therefore turning to III-V materials, which can be combined in heterostructures. Furthermore, lineTFET configurations are being investigated in which the tunneling is oriented more orthogonal to the gate than in the standard pointTFET configurations.

Standard commercial semiclassical modeling approaches are poorly suited to assess TFET performance for these new material systems and configurations. This is because they neglect quantum phenomena such as size-induced and field-induced quantum confinement, and reflections at the heterojunction.

In this thesis, we therefore develop a fully quantum mechanical simulator, called Pharos, based on the multi-band envelope function formalism to simulate BTBT in direct semiconductors. Our approach allows for computationally efficient performance predictions and optimization of heterostructure TFETs, and enables the comparison between different III-V material options and configurations.

We implement our formalism for a two-, fifteen-, and thirty-band model, with each subsequent model enabling the simulation of a wider variety of configurations. The two-band model is only suited to simulate pointTFETs. For this configuration, we find a counteracting effect between gate control

and size-induced quantum confinement for decreasing device dimensions. The fifteen-band model is implemented with a spectral approach and enables the simulation of lineTFETs, which we compare to pointTFET configurations. We find that an optimized pocketed pointTFET has similar performance than an optimized pocketed lineTFET. We also introduce an improved source design, which brings the performance of the pTFET to the same level as the nTFET, enabling complementary logic applications. With a thirty-band model, we assess whether strain can further improve the heterostructure TFET performance. We find that uniform strain can improve the ON-currents of heterostructure TFETs, if our improved source design is applied. We also assess non-uniform strain profiles which arise at a lattice-mismatched heterojunction and find that the lattice mismatch can be used as an additional design parameter to enlarge the TFET design space.

We also develop a self-consistent procedure, which couples the calculated charge density to the electrostatic potential using a Gummel scheme for Poisson's equation. This allows us to identify the impact of quantum effects on the electrostatic potential. We find that self-consistent simulations are required for strongly confined structures.

Finally, we report on simulations done during a research stay at Purdue University in IN, USA, which compare the sensitivity to electron-phonon scattering in conventional heterostructure TFET and resonant TFET configurations. We find a larger sensitivity for the resonant TFET, although it still offers superior performance to the conventional configuration.

Beknopte samenvatting

De schaling van de metaal-oxide-halfgeleider-veldeffecttransistor (MOSFET) is sinds de jaren 1960 de drijvende kracht achter de enorme toename in computationele kracht van alledaagse electronica. Tegenwoordig bereikt deze trend zijn limieten, omdat de voedingsspanning van de MOSFET niet meer op hetzelfde tempo kan geschaald worden als de componentafmetingen. Dit komt door een onderlimiet op de subthreshold swing (SS). Het gevolg is dat de vermogendichtheid van geïntegreerde circuits onaanvaardbaar toeneemt met elke nieuwe generatie.

De tunnelveldeffecttransistor (TFET) is bedacht om deze nefaste evolutie te doorbreken. Het werkingsmechanisme gebaseerd op band-tot-bandtunneling (BTBT) laat een lage SS toe, en maakt zo een lage voedingsspanning mogelijk. Implementaties in silicium hebben echter onvoldoende AAN-stroom getoond. TFET-onderzoek gaat daarom steeds meer in de richting van III-V-materialen, die gecombineerd kunnen worden in een heterostructuur. Daarnaast worden ook lijnTFET-configuraties onderzocht, waarin de tunneling meer orthogonaal gericht is op de gate dan voor de standaard puntTFET-configuratie.

Standaard commerciële semiklassieke modelleringstechnieken zijn niet geschikt om de TFET-performantie in te schatten voor deze nieuwe materiaalsystemen en configuraties. Ze verwaarlozen namelijk belangrijke kwantummechanische effecten, zoals veld-en grootte-geïnduceerde kwantumopsluiting, en reflecties aan de heterojunctie.

In deze thesis ontwikkelen we daarom een volledig kwantummechanische simulator, Pharos genaamd, voor het simuleren van BTBT in directe halfgeleiders, gebaseerd op het omhullendefunctieformalisme. Onze aanpak laat toe om op een computationeel efficiënte manier de performantie te voorspellen en optimalisaties uit te voeren van heterostructuur-TFETs. Zo kunnen de verschillende III-V-materiaalopties en configuraties vergeleken worden.

We implementeren ons formalisme voor een twee-, vijftien-, en dertigbandenmodel, waarbij elk opeenvolgend model toelaat een bredere waaier van configuraties te onderzoeken. Het tweebandenmodel laat enkel de simulatie van puntTFETs toe. Voor deze configuratie vinden we een tegenwerkend effect tussen de gatecontrole and de grootte-geïnduceerde kwantumopsluiting bij afnemende componentafmetingen. Het vijftienbandenmodel wordt geïmplementeerd met een spectrale methode en laat ook de simulatie van lijnTFETs toe, die we vergelijken met puntTFETs. We stellen vast dat een geoptimaliseerde puntTFET met pocket een gelijkaardige performantie vertoont als een lijnTFET met pocket. We introduceren ook een verbeterd sourceontwerp dat de performantie van de pTFET op het niveau van de nTFET brengt en zo complementaire logica mogelijk maakt. Met het dertigbandenmoden, ten slotte, onderzoeken we of rek de performantie van heterostructuur-TFETs verder kan verbeteren. We ontdekken dat uniforme rek de AAN-stroom kan verbeteren, mits de toepassing van ons verbeterde sourceontwerp. We bestuderen ook niet-uniforme rekprofielen die het gevolg zijn van een verschil in roosterconstante aan de heterojunctie en stellen vast dat de relatieve verhouding van de roosterconstantes kan gebruikt worden als een extra ontwerpparameter.

We ontwikkelen ook een zelfconsistente procedure die de berekende ladingsdichtheid koppelt aan de elektrostatische potentiaal met de Gummelmethode voor de Poissonvergelijking. Dit laat ons toe de impact van kwantumeffecten op de potentiaal in te schatten. We stellen vast dat zelfconsistente simulaties onontbeerlijk zijn voor sterk opgesloten structuren.

Ten slotte presenteren we een verslag van simulaties die uitgevoerd zijn tijdens een onderzoeksverblijf aan Purdue University in IN, VS, met als doel het vergelijken van de gevoeligheid voor elektron-fononverstrooiing van conventionele TFETs en resonante TFETs. We vinden een grotere gevoeligheid voor de resonante TFET, al behoudt die een betere performantie dan de conventionele configuratie.

Glossary

Acronyms

BGN	BandGap Narrowing
BTBT	Band-To-Band-Tunneling
CMOS	Complementary Metal-Oxide-Semiconductor
DG	Double Gate
DOS	Density Of States
EF	Envelope Function
EOT	Equivalent Oxide Thickness
FD	Finite Difference
FE	Finite Element
FIQC	Field-Induced Quantum Confinement
GAA	Gate-All-Around
GWF	Gate Work Function
LDOS	Local DOS
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
NEGF	Non-Equilibrium Green's Function
NEMO5	NanoElectronics MOdeling package, version 5
NPO	Non-Polar Optical
POP	Polar Optical Phonon
QM	Quantum Mechanical
SC	SemiClassical
SIQC	Size-Induced Quantum Confinement
SG	Single Gate
SRAM	Static Random Access Memory
SRH	Shockley-Read-Hall
SS	Subthreshold Swing
SUR	Successive UnderRelaxation
TAT	Trap-Assisted Tunneling
TB	Tight-Binding

TFET	Tunnel Field-Effect Transistor
WF	Wave Function approach
WKB	Wentzel-Kramers-Brillouin

Mathematical symbols

$a_\mathrm{c}, b_\mathrm{v}, d_\mathrm{v}$	Deformation potentials
C_{11}, C_{12}, C_{44}	Elements of the stiffness matrix
C_GD	Gate-drain capacitance
C_GS	Gate-source capacitance
C_L	Load capacitance
C_M	Miller capacitance
C_ox	Oxide capacitance
D_ac	Deformation potential constant
D_opt	Optical coupling constant
E	Energy
$E_\mathrm{c,v}$	Band edge energy of conduction, valence band
E_EA	Electron affinity
$E_\mathrm{Fn,Fp}$	Quasi Fermi level of electrons, holes
E_G	Bandgap
$E_\mathrm{G,eff}$	Effective bandgap at a heterojunction
F	Electric field
F_n	Envelope function corresponding to band n
$f_\mathrm{S,D}$	Fermi-Dirac distribution of source, drain contact
G	Generation rate
$G^{R,<,>}$	Retarded, lesser, greater Green's function
h	Planck constant
$\hbar$	Reduced Planck constant
H_0	Hamiltonian of the closed system
H_{nm}	Hamiltonian matrix element of bands n and m
H_S	Strain Hamiltonian
I_DS	Drain-source current
I_ON	ON-state current
I_OFF	OFF-state current
k_B	Boltzmann constant
$k_{x,y,z}$	Wave number in x, y, z-direction
L_source	Length of source region
L_channel	Length of channel region
L_drain	Length of drain region
L_gate	Gate length
L_gs	Gate-source overlap

L_{ud}	Gate-source underlap
m_e	Free electron mass
$m^*_{\mathrm{c,v}}$	Conduction, valence band effective mass
m_{R}	Reduced effective mass
n	Electron density
n_0	Bose-Einstein distribution
N	Number of bands
N_{po}	Pocket doping
N_{source}	Source doping
$N_{x,z}$	Number of discretization points in the x,z-direction
N_{k_z}	Number of spectral components
p	Hole density
$\mathbf{p}_{nm}$	Interband momentum matrix element of bands n and m
q_{e}	Elementary charge
$S^{l\to k}$	Basis function transformation of layer l to layer k
T	Temperature
T_{body}	Body thickness
T_{i}	Width of the intrinsic region of a p-i-n diode
$T_{\mathrm{Kane,WKB}}$	Transmission probability in the Kane, WKB model
T_{po}	Pocket thickness
U_n	Basis function corresponding to band n
V	Volume
V_{c}	Crystal potential energy
V_{DD}	Supply voltage
V_{e}	External applied potential energy
V_{DS}	Drain-source voltage
V_{GS}	Gate-source voltage
V_{th}	Threshold voltage
α,β	Orbital and atom indices
ϵ_s	Permittivity (dielectric constant) of the semiconductor
$\overline{\overline{\varepsilon}}$	Strain tensor
ε_{ij}	Element of the strain tensor
θ_l	Logical step function
ρ	Material density
$\Sigma^{R,<,>}_{\mathrm{S/D}}$	Retarded, lesser, greater self-energy of source and drain
$\Sigma^{R,<,>}_{\mathrm{el\text{-}ph}}$	Retarded, lesser, greater self-energy of electron-phonon interaction
σ_{ij}	Element of the stress tensor
ϕ_n	Löwdin orbital with quantum number n
$\chi_{n\alpha}$	Envelope function of subband α of band n
ψ	One-electron wave function
ω_0	Phonon frequency

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the fundamental endeavors of engineering science is to harness unique physical phenomena to develop devices that ultimately benefit humankind's technological development. A key aspect in this process is the ability to predict the interplay between such phenomena and the design of a device. In this thesis, we consider a particular electronic device: the tunnel field-effect transistor (TFET). The TFET employs the quantum mechanical (QM) phenomenon of band-to-band tunneling (BTBT) to enable low-power switching in integrated circuits. This thesis outlines a QM approach to predict how the TFET operation is influenced for different choices of materials and configurations, with the aim of optimizing its performance. The approach is implemented as a simulator called Pharos.

This chapter serves as a general introduction, starting with a sketch of the state of the field at the beginning of the PhD project (Section 1.1), followed by the goals and organization of the thesis (Section 1.2). The TFET itself is introduced in the next chapter by means of an article published in the *Wiley Encyclopedia of Electrical and Electronics Engineering*.

1.1 The beginning: state of the field

At the outset of this PhD project in 2012, research on the TFET was at a crossroads. Newly developed QM methods at imec to simulate group IV TFETs had tempered the enthusiasm fueled by initial over-optimistic semiclassical (SC) simulations [1]. At the same time, experimental realizations seemed to

confirm that silicon (Si) TFETs would not achieve ON-currents (I_{ON}) sufficient to replace the conventional metal-oxide-semiconductor field-effect transistor (MOSFET) in both low-voltage and high performance applications [2, 3, 4].

Once heralded as the surefire low-voltage successor to the MOSFET thanks to its steep switching characteristics, the TFET was now forced to turn to less conventional material systems. Among these were other group IV materials such as (strained) germanium (Ge), but also compound materials from group III and V, the so-called "III-V materials" [5, 6, 7, 8]. This class of materials features a wide choice in bandgaps, which is further enlarged by the possibility to combine them in a heterostructure [9, 10]. III-V materials were also being considered for MOSFET applications, because they offer high carrier mobilities. Would one or a combination of these III-V materials also have the ideal properties for TFET?

While TFET research was starting to leave the path of conventional materials like Si, it was also looking for ways to improve the device configuration itself. The standard TFET configuration, also called a "pointTFET", basically a gated p-i-n diode, was facing competition from a new concept with a different tunneling orientation: the "lineTFET" [11, 12]. In the lineTFET, tunneling is oriented perpendicular to the gate, which should result in a more abrupt switching. Both the pointTFET and the lineTFET also saw the introduction of localized regions of doping, called "pockets", to boost I_{ON} or reduce unwanted quantum confinement effects. For group IV materials, simulations with the new QM methods of both configurations were giving the advantage in terms of performance to the lineTFET [13, 14], both with and without pockets. Would this advantage also hold in a direct bandgap III-V material system?

Preliminary simulation results were showing another path to improve both group IV and III-V TFETs: mechanical strain [15, 16, 17]. Commonly applied in MOSFETs to boost the carrier mobility, the band warping effect of strain holds the promise of changing the tunneling properties of a material at will. Obtaining the optimal TFET performance would then just be a matter of applying an appropriate amount of stress to the material. However, these initial simulations also showed that strain changes multiple properties of the TFET at the same time, not all beneficial. Predicting the effect on the overall performance of a given strain configuration was therefore found to be non-trivial. This was particularly important for heterostructures of materials with different lattice constants: such configurations show a complicated, non-uniform strain profile around the junction. Could some of these strain configurations boost the III-V TFET performance, and if so, which ones?

To answer these questions on material system, device configuration and strain impact, previous experience with group IV materials had made it clear that SC simulations could not be trusted and that therefore a QM approach was in order.

This intuition was further confirmed by an initial 1D QM study of tunneling at a heterostructure junction, which showed reflections at the interface that are completely disregarded in SC simulations [18]. Some QM simulation packages capable of simulating III-V TFETs did exist, but were too computationally expensive to allow the exploration of the vast design space of materials and configurations [19]. Hence, a clear need existed for a computationally efficient QM simulator for III-V heterostructure TFETs.

1.2 Goals and organization of the thesis

1.2.1 Goals

From the questions posed in the previous section, we distill the following key goals and subgoals of this PhD project:

1. **Develop a QM formalism to simulate transport in heterostructure TFETs in a computationally efficient way.**
   - Select and/or derive a QM formalism to simulate band-to-band-tunneling for 2D potentials in direct bandgap materials.
   - Enable the simulation of heterostructure device configurations.
   - Implement the QM formalism numerically such that it can be used for configuration optimization. Simulation time should therefore be in the range of hours or days, not weeks.

2. **Assess the viability of direct bandgap III-V TFET.**
   - Determine whether the lineTFET concept works in a direct bandgap III-V material system.
   - If the previous item has a positive outcome: determine which configuration has the most promising transfer characteristics in a III-V material system: lineTFET or pointTFET.
   - Assess the complementarity of III-V TFET.

3. **Investigate whether strain can be a performance booster for III-V TFETs.**
   - Investigate whether uniform strain improves heterostructure TFET performance.
   - Investigate the impact of non-uniform strain at lattice-mismatched heterojunctions and whether it can be used as a performance booster.

1.2.2 Organization

After the introduction of the TFET operation, device physics and configuration options in Chapter 2, the basic formalism for our QM simulation approach is selected and derived for our application in Chapter 3.

The following three chapters have a similar structure. They discuss implementations of the basic formalism for band structure models with an increasing number of bands. For each model, we explain why we need that number of bands in terms of the TFET configurations that we want to investigate. We discuss the selected bands, the challenges faced to implement the model and solutions to these challenges, which often require an extension of the basic formalism. We then use the implemented formalism to simulate various TFET configurations to find answers to the questions posed in the previous section.

The three discussed band models are the two-band, fifteen-band and thirty-band models. Chapter 4 starts off with the simplest model, containing only two bands. With this model, we investigate the effects of confinement on heterostructure TFET performance. Limitations of the two-band model in simulating lineTFET configurations lead to the extension to fifteen bands in Chapter 5. For the implementation of the fifteen-band model, the basic formalism is significantly modified with a spectral approach. With the fifteen-band implementation, we optimize and compare pocketed III-V lineTFET and pointTFET configurations. We also introduce an improved source design to improve the performance of III-V pTFET and enable complementary circuit implementations. To simulate the effects of strain, we turn to a thirty-band model in Chapter 6, which includes the spin-orbit interaction. We discuss the effects of uniform and non-uniform strain profiles on heterostructure TFET performance.

In Chapter 7, we present a self-consistent simulation procedure that expands the application domain of our simulator to strongly confined device configurations. With the self-consistent procedure, we also check some of the non-self-consistent results of the previous chapters.

Finally, Chapter 8 is somewhat distinct from the other chapters, as it reports work carried out during a three-month research stay at Purdue University and is not based on the QM approach developed throughout the rest of the thesis. Instead, this chapter investigates the effect of electron-phonon scattering on the performance of a conventional and a resonant TFET with the QM simulation package NEMO5. The thesis concludes with the major findings and future prospects in Chapter 9.

Chapter 2

The tunnel field-effect transistor

After the general introduction of the previous chapter, this chapter introduces the TFET: its history, operation principles, modeling and implementation options. The chapter is in fact a slightly abridged version of an article published in the *Wiley Encyclopedia of Electrical and Electronics Engineering* in November 2016, authored by Devin Verreck, Anne S. Verhulst and Guido Groeseneken. The article was written during the last year of this PhD project and therefore reflects the state-of-the-art at this time. Sections discussing topics and results that are discussed in detail in the other chapters of the thesis have been omitted. For an overview of the state-of-the-art at the beginning of this PhD project, we refer to Section 1.1.

Contributions of the first author:

- Literature study
- Writing of text
- Creation of figures

2.1 Introduction

The tunnel field-effect transistor (TFET) is a semiconductor device aimed at low-power logic applications that employs band-to-band tunneling (BTBT) as a carrier injection mechanism to obtain a subthermionic subthreshold swing (SS). In particular, it relies on the energetic filtering of the tail of the electron Fermi-Dirac distribution to go below the fundamental 60 mV/dec SS limit at room temperature of a metal-oxide-semiconductor FET (MOSFET). The goal is to combine a low leakage current with a low SS to allow the TFET to be more energy efficient than a MOSFET.

Research in TFET has been driven by the fundamental power issues encountered by MOSFET as device scaling continues along the path of Moore's law. This law, in its most common formulation today, states that the number of electronic components per chip resulting in a minimum component cost, doubles approximately every two years [20, 21]. Also, the rise of mobile applications and the Internet of Things, which contain a plethora of always-on sensor nodes, has increased the demand for devices with low supply voltage (V_{DD}) and low-leakage operation [22, 23]. The TFET aims to fulfill this demand by exploiting the quantum mechanical phenomenon of tunneling. Although it is a leakage mechanism for MOSFET in today's scaled architectures, tunneling enables the TFET to go beyond the inherent ON-current (I_{ON}) - OFF-current (I_{OFF}) trade-off that hampers the low-power performance of the MOSFET. Several challenges still remain, however, before TFET can be implemented as a low-power replacement for MOSFET.

Several excellent overview works exist which summarize the TFET state-of-the-art or go into more detail on specific TFET topics [24, 25, 26, 27]. This article aims to give a broad overview of the TFET field, with a distinct focus on device physics and architecture options. For more details, the reader is invited to consult the references included in each section. The article is structured as follows. First, the power issue of the MOSFET, which the TFET seeks to solve, is outlined in Section 2.2. Next, the basic operation of the TFET is explained in Section 2.3, along with the two main types of tunneling in Section 2.4. This is followed by an overview of different approaches to model TFET operation in Section 2.5. The main performance challenges for TFET are then presented in Section 2.6. In Section 2.7, different material options for TFET are considered. The subsequent sections are devoted to various implementation options that are being researched to improve TFET performance, such as dopant pockets (Section 2.8), specific gate configurations (Section 2.9) and strain (Section 2.10). Section 2.11 discusses attention points upon using the TFET in a circuit. Finally, Section 2.12 gives a brief update on experimental work in literature. Section 2.13 concludes the article and provides a future outlook.

2.2 MOSFET power issue

The inherent trade-off between I_{ON} and I_{OFF} for decreasing V_{DD} lies at the heart of the power density issue of MOSFET-based logic. This can be understood by looking at the total dissipated power of a circuit of MOSFET-based logic gates, which consists of a static and a dynamic component [28]:

$$P_{\mathrm{tot}} = P_{\mathrm{static}} + P_{\mathrm{dynamic}} = N_{\mathrm{g}} I_{\mathrm{OFF}} V_{\mathrm{DD}} + \alpha C_{\mathrm{tot}} V_{\mathrm{DD}}^2 f \tag{2.1}$$

with N_{g} the amount of gates, α the fraction of active gates, C_{tot} the total load capacitance of all gates and f the switching frequency. Based on Eq. (2.1), a key element in the so-called Dennard scaling of MOSFET, proposed in 1974 [29], is the reduction of V_{DD} as the physical transistor dimensions are decreased with every new generation in order to maintain a constant power density (note that although C_{tot} decreases, the V_{DD} reduction also results in an increase of f). The threshold voltage (V_{th}) is decreased accordingly to maintain sufficient I_{ON}, which is proportional to (V_{DD} - V_{th})^x. The Dennard scaling paradigm has enabled the continuation of Moore's law until the end of the 20th century. However, Dennard scaling eventually leads to an untenable increase in the static power component, as the OFF-state leakage is exponentially dependent on V_{th} [24]:

$$I_{\mathrm{OFF}} \sim e^{\frac{-V_{\mathrm{th}}}{nkT/q}} \tag{2.2}$$

where kT/q is the thermal voltage with k the Boltzmann constant, T the temperature and q the elementary charge, and where n is the body factor, equal to $\left(1 + \frac{C_{\mathrm{d}}}{C_{\mathrm{ox}}}\right)$ with C_{d} and C_{ox} respectively the depletion and oxide capacitance of a planar MOSFET. Around the year 2002, the path of Dennard scaling was therefore abandoned, with the scaling of V_{DD} slowing down with respect to the scaling of the physical transistor dimensions. As a result, power density has been increasing as more and more transistors are included on a chip, with each transistor consuming roughly the same power as the previous generation. Notably, a significant portion of this power is consumed in the OFF-state. The rising power density leads to issues with cooling and reliability.

The origin of the exponential dependence in Eq. (2.2) is the Fermi-Dirac distribution of the charge carriers in the source region. In the subthreshold regime, also called weak inversion, a potential energy barrier in the MOSFET channel region prevents low energy carriers in the source from flowing to the drain contact. However, the high energy carriers in the exponential tail of the distribution can still diffuse over the barrier in a process called thermionic emission. This leads to an exponential dependence of the drain-source current I_{DS} on the gate-source voltage V_{GS} [30]:

$$I_{\mathrm{DS}} \approx I_{\mathrm{D0}} e^{\frac{V_{\mathrm{GS}} - V_{\mathrm{th}}}{nkT/q}} \quad \text{for } \mathrm{V_{GS}} < \mathrm{V_{th}} \tag{2.3}$$

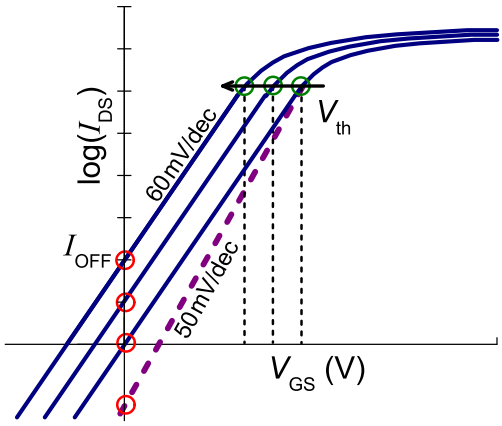

Figure 2.1: Schematic transfer characteristics of a MOSFET with an ideal SS and a steep slope device with a sub-60 mV/dec SS (dashed), illustrating that the SS determines I_{OFF}. Scaling of the threshold voltage V_{th} increases the OFF-current I_{OFF} exponentially.

with I_{D0} the current at $V_{\mathrm{GS}} = V_{\mathrm{th}}$.

The inverse of the slope of Eq. (2.3) determines the trade-off between I_{ON} and I_{OFF}. It is called the subthreshold swing (SS) and is used as a metric for the switching steepness of the device. SS is defined as the increase in V_{GS} that is required to increase I_{DS} with one order of magnitude [24, 29], and can be expressed as:

$$SS = \frac{V_{\mathrm{GS}}}{\log_{10}(I_{\mathrm{DS}})} \approx \left(1 + \frac{C_{\mathrm{d}}}{C_{\mathrm{ox}}}\right) \frac{kT}{q} \ln(10). \tag{2.4}$$

At room temperature ($T = 300$ K), the SS of a MOSFET is therefore theoretically limited to about 60 mV/dec. In actual implementations, non-idealities result in SS values which can be significantly higher. Fig. 2.1 illustrates graphically that the value of the SS determines the intersection with the $V_{\mathrm{GS}} = 0$ V axis, which corresponds to I_{OFF}. I_{OFF} increases exponentially as V_{th} is decreased and the curve shifts leftwards. As I_{OFF} becomes unacceptably high, it prevents further concurrent scaling of V_{DD} and V_{th} and hence leads to aforementioned power density issues in highly scaled technologies. The underlimit on the SS makes this a fundamental trade-off.

Several transistor concepts have been proposed to break the I_{ON}-I_{OFF} trade-off with a SS lower than the MOSFET limit. Examples include concepts which use negative capacitance [31], impact-ionization [32] and mechanical switches [33].

However, these concepts give rise to hysteretic behavior and typically require a high operating voltage (> 1 V) at one of the transistor contacts.

The TFET is an alternative transistor concept, compatible with CMOS technology, proposed to be capable of a SS lower than the MOSFET limit. This is possible because the TFET relies on quantum mechanical BTBT instead of thermionic emission as the carrier injection mechanism. The basic TFET structure was first proposed in 1978 as a "surface channel tunnel junction", aimed at investigating subband splitting and many-body effects in quasi-2D systems [34]. However, it was not until 1987 that BTBT was suggested as the working principle of a transistor in a DRAM trench transistor cell [35]. Major research efforts started after 2004, when a carbon nanotube TFET was demonstrated with a sub-60 mV/dec SS at room temperature [36] and the use of heterostructures was conceptually introduced [37]. Since then, significant research efforts have been invested worldwide in improving the TFET performance by optimization of the device architecture and material system.

2.3 TFET basics

2.3.1 Basic structure and operating principle

The basic TFET structure is similar to that of a MOSFET: it contains two contact regions and an intrinsic or lowly doped channel region, covered by a gate dielectric and a gate contact (see Fig. 2.2(a)). In contrast to a MOSFET, however, the TFET contact regions have an opposite doping polarity, resulting in a p-i-n profile. In an nTFET, the p-type region acts as the source region, while in a pTFET, the n-type region acts as the source. Many variations on the basic configuration are possible, with different gate overlaps or doping profiles. These will be discussed in Sections 2.8 to 2.10. First, the working principle will be explained for a basic p-i-n nTFET in a semiclassical picture. In this discussion, it is assumed that the doping is such that the Fermi-level in the source is aligned with the valence band edge.

The TFET operates by enabling and preventing BTBT between the source and the channel region by modulation of the electrostatic potential in these regions with the gate contact (see Figs. 2.2(b) and (c)). The source and drain contacts are biased such that the p-i-n diode is in reverse bias. In the OFF-state, the only current that flows is the reverse leakage current of the p-i-n diode. This leakage current is typically caused by minority carrier diffusion currents and by defect-assisted processes, such as Shockley-Read-Hall (SRH) generation

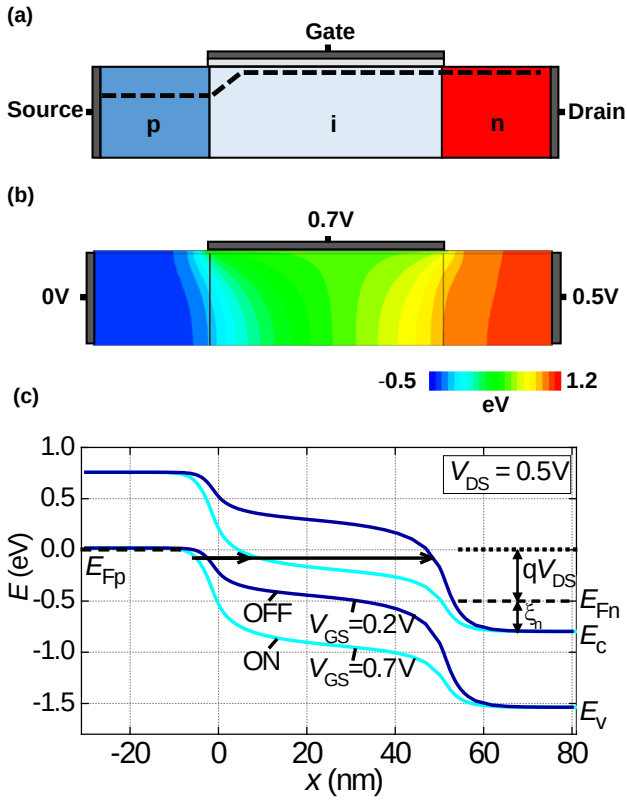

Figure 2.2: (a) Basic p-i-n TFET configuration. (b) 2D electrostatic profile of a p-i-n $In_{0.53}Ga_{0.47}As$ TFET in the ON-state. (c) Energy band diagram along the cutline in (a) in the OFF (dark lines) and ON-state (light lines). The black arrows illustrate the shortening of the available tunnel paths. The quasi-Fermi level for the holes in the source (E_{Fp}) coincides with the valence band edge, while at the drain side, there is a degeneracy ξ_{n}.

and trap-assisted-tunneling (TAT) (see Section 2.6.3). As V_{GS} is increased, the source is depleted and the electric field at the tunnel junction rises. This corresponds to an increasing band bending at the source-channel junction. At a given V_{GS}, called the onset voltage (V_{onset}), the conduction band in the channel crosses over with the valence band in the source, such that tunneling transitions between these two bands become available. These transitions can be direct, between the maximum of the valence band and the minimum of the conduction band at the Γ-point, or indirect, between the maximum of the valence band and the minimum of one of the conduction band valleys. In the indirect case, the transition is assisted by a phonon. The tunneling transitions, whether direct or indirect, form the current generating process of the TFET.

2.3.2 Transfer characteristics

Fig. 2.3 plots an example of typical TFET transfer characteristics, I_{DS} as a function of V_{GS}, showing the different operating regimes. The tunneling current rises as V_{GS} is increased above V_{onset}, since the length of the available tunnel paths decreases and the electron tunneling probability is exponentially dependent on the tunneling distance. At V_{GS} equal to V_{DD}, the transistor is in the ON-state, and I_{DS} at this point is I_{ON}. The transition from the OFF to the ON-state is, in analogy to the MOSFET, called the subthreshold swing (SS), although there is no fixed-swing subthreshold regime as in a MOSFET (see Section 2.6.2), and hence also no V_{th}. The I-V curve in Fig. 2.3 shows a SS lower than the 60 mV/dec MOSFET limit.

The low SS of the TFET originates from the energetic filtering effect of BTBT carrier injection. Fig. 2.4(a) illustrates how the bandgap of the source material cuts off a significant part of the exponential tail of the Fermi-Dirac distribution. This band-pass filter action effectively cools the carrier distribution. The origin of the SS-limit in a MOSFET, which was discussed in Section 2.2, is thus removed. The TFET is therefore able to obtain a SS lower than 60 mV/dec at room temperature. When the Fermi-level in the source is no longer aligned with the valence band in the case of a highly doped source, the SS can be degraded if the source degeneracy is too large. Figs. 2.4(b) and (c) show that a part of the exponential tail is then no longer filtered. If this part of the tail results in observable current, it deteriorates the SS. On the other hand, Figs. 2.4(b) and (c) also show that a higher source doping increases the electric field at the tunnel junction, which leads to a higher I_{ON}. This SS-I_{ON} trade-off, along with other parameters influencing SS, is discussed further in Section 2.6.2.

The TFET is an ambipolar device, meaning that an nTFET operates as a pTFET when a negative V_{GS} is applied. In this case, the depletion occurs in

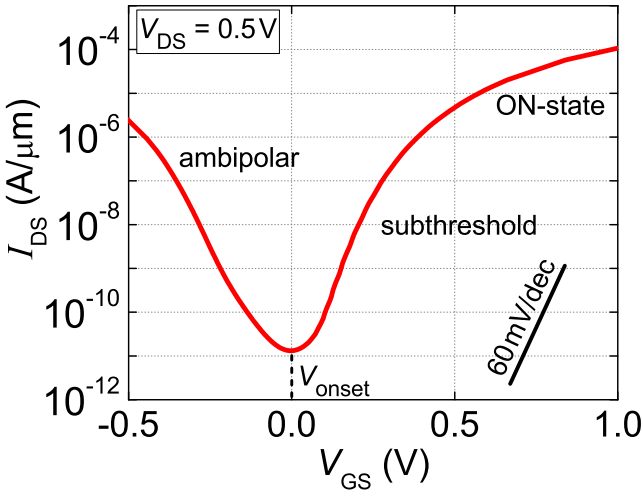

Figure 2.3: Semiclassical (SC) simulated transfer characteristics of an $In_{0.53}Ga_{0.47}As$ p-i-n TFET as shown in Fig. 2.2(a). The TFET body thickness (T_{body}) is 20 nm with a source doping of $5x10^{19}cm^{-3}$ and an EOT of 0.6 nm.

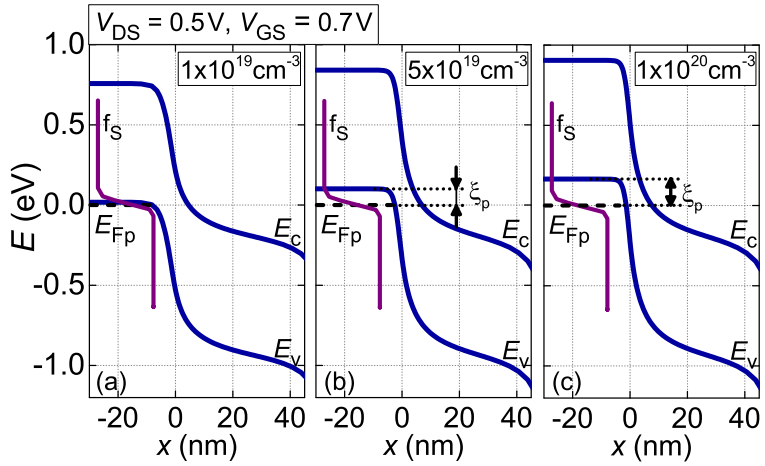

Figure 2.4: Energy band diagram of the p-i-n $In_{0.53}Ga_{0.47}As$ TFET of Fig. 2.2 with different degrees of source doping degeneracy ξ_p, ranging from (a) no degeneracy to (b) medium degeneracy to (c) high source degeneracy. As the source degeneracy increases, less of the exponential tail of the Fermi-Dirac distribution f_S is filtered by the source bandgap. For the ease of extraction, a cross-section parallel to the gate has been used.

the n-type drain region as the electric field rises at the channel-drain junction. After cross-over of the valence band in the channel with the conduction band in the drain, a hole tunneling current is injected into the channel and the current increases with decreasing V_{GS} (see Fig. 2.3). The ambipolar behavior can be both an advantage and a disadvantage, and will be further discussed in Section 2.11.1.

2.3.3 Output characteristics

The source doping, drain doping and V_{DS} determine the energetic window available for tunneling (see Fig. 2.2(b)). V_{DS} fixes the hole quasi-Fermi level E_{Fp} in the source relative to the electron quasi-Fermi level in the drain E_{Fn}. The total tunneling window, the energetic distance between the valence band in the source and the conduction band in the drain, is then the sum of (E_{Fp}-E_{Fn}) and any doping degeneracies in source ξ_p and drain ξ_n (see Fig. 2.4). For a constant V_{GS} and starting at a V_{DS} of 0 V, an increase in V_{DS} enlarges the tunneling window and increases I_{DS} (see Fig. 2.5). The increase of I_{DS} with V_{DS} continues until the conduction band edge in the drain falls below the conduction

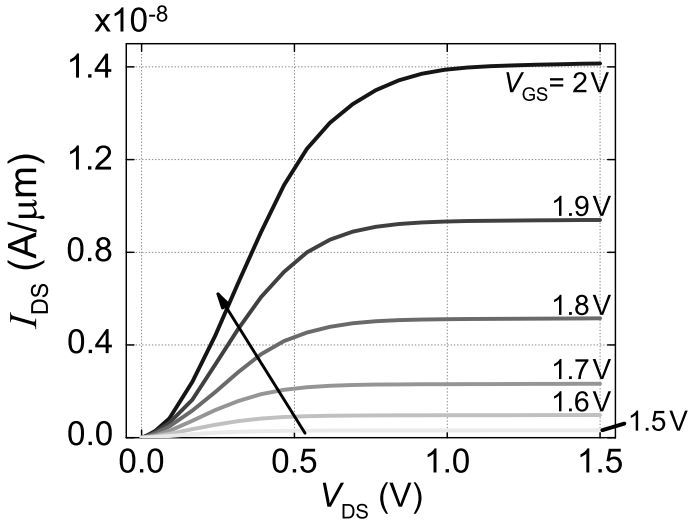

Figure 2.5: SC simulated output characteristics of a Si p-i-n TFET as shown in Fig. 2.2(a) for varying V_{GS}. The TFET T_{body} is 20 nm with a source doping of $5x10^{19}cm^{-3}$ and an EOT of 0.6 nm.

band edge in the channel. Beyond this V_{DS}, called $V_{DS,SAT}$, I_{DS} saturates, since an increase in V_{DS} no longer has an impact on the tunneling window. Since the conduction band edge in the channel is determined by V_{GS}, $V_{DS,SAT}$ increases with increasing V_{GS}, just like in a MOSFET.

Fig. 2.5 also shows that in contrast to the MOSFET, the onset of the TFET output characteristics can be superlinear. This occurs when the BTBT is inefficient for small V_{DS}, e.g. as a result of a large tunnel path caused by a large equivalent oxide thickness (EOT), a low source doping or as a result of very low carrier density available for BTBT [38, 39, 40] (see also Section 2.6.1). In the superlinear regime, the output current is small. This is unwanted, since it increases the settling time of a TFET inverter. Further circuit implications of this superlinear onset will be discussed in Section 2.11.4.

2.4 Point versus line tunneling

After introducing the basic operating principle in the previous section, a first major distinction in tunneling types can now be made. Based on the alignment of the electric field induced by the gate with the dominant tunnel paths, two types of tunneling can be defined in a TFET: point tunneling (Section 2.4.1) and line tunneling (Section 2.4.2), which can both be present in the same device (Section 2.4.3).

2.4.1 Point tunneling

Point tunneling is the dominant type in a standard p-i-n/n-i-p TFET as depicted in Fig. 2.2(a) or Fig. 2.6(a), with the gate predominantly covering the channel. This configuration is also called a pointTFET. The term "point tunneling" originates from the assumption that the tunnel paths curve around a central point at the interface between gate metal and gate dielectric, right above the tunnel junction. In this approximate semiclassical view, the tunnel paths are assumed to lie on circles formed by the electric field lines between source and gate, starting at the tunnel junction and ending at the interface between gate-dielectric and channel [39]. Note that the circular form of the field lines is based on the assumption of infinitely high source doping. At onset, only the longest tunnel paths are available, which then gradually shorten as V_{GS} is increased toward the ON-state. In Fig. 2.2(a), these circular lines have been replaced with a straight line indicating the shortest allowed tunnel path, with roughly same start and end point as the circular lines.

The pointTFET performance is influenced by field-induced quantum confinement (FIQC), which is present in the triangular well formed by the conduction band edge in the channel and the dielectric barrier [41] (see Fig. 2.7, although for a lineTFET, the shape of the well is similar). In this well, subbands are formed. The onset of BTBT is thereby shifted to a higher V_{GS}, since the band edge determining the allowed tunnel path inside the well is defined by the first subband level, which is higher in energy than the bulk value of the band edge. As V_{GS} is increased and the triangular well deepens, the first subband level shifts down at a slower pace than the bulk band edge. Hence, the tunneling window increases more slowly than it would in the absence of FIQC. This has a stretching effect on the transfer characteristics, and hence negatively impacts SS. The degree of this confinement is determined by the effective mass of the band structure valley to which the tunneling transition is taking place and therefore depends on the material choice for the channel. The impact of FIQC can be alleviated with the introduction of a doping pocket at the source-channel interface, which forces the tunnel paths more parallel to the gate (see Section 2.8.1).

2.4.2 Line tunneling

Line tunneling can be induced with a large overlap of the gate over the source (see Fig. 2.6(b)). Such a configuration is also called a lineTFET. In contrast to point tunneling, the tunnel paths are equally long parallel straight lines perpendicular to the gate dielectric in the ideal case of no parasitic paths. For an increasing V_{GS}, the energy bands bend toward the gate dielectric until cross-over occurs between the conduction band edge at the gate dielectric and the valence band edge in the bulk, such that BTBT becomes possible (see Fig. 2.7). In the idealized case of a uniform field underneath the gate dielectric, implying that also the drain voltage impact on the source region is completely neglected, the onset is more abrupt than for point tunneling. All tunnel paths underneath the gate-source overlap become available at the same amount of band bending, after which they shorten uniformly for increasing V_{GS}. The tunneling is also located closer to the gate than in a pointTFET, while the electric field is in line with the tunnel paths. This means a smaller increase in V_{GS} is required to achieve a given amount of band bending. Additionally, in the ON-state, the band bending is stronger, resulting in shorter tunnel paths. Important to note is that I_{ON} is proportional to the gate-source overlap.

The lineTFET's performance is more heavily impacted by FIQC than the pointTFET, since the triangular well underneath the gate is more pronounced. The stretching effect due to the slower increase in the tunneling window, also seen in the pointTFET, can therefore have a negative impact on the SS. Additionally,

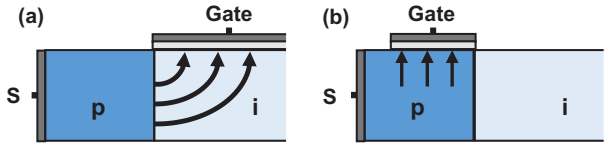

Figure 2.6: Source-channel region of (a) a point tunneling and (b) a line tunneling TFET configuration. The arrows schematically indicate the tunneling paths.

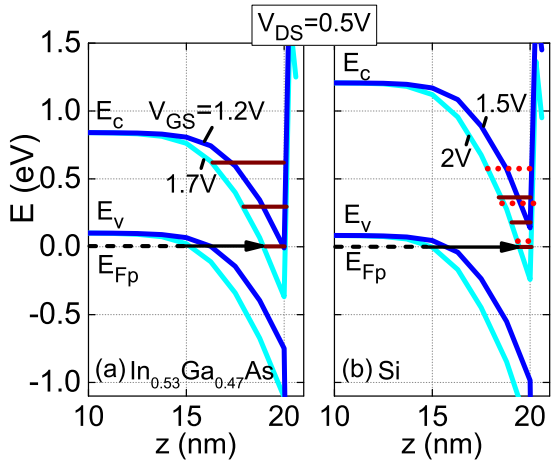

Figure 2.7: Energy band diagrams of a lineTFET in the OFF (dark lines) and ON-state (light lines) along a vertical cutline through the center of the gate in the [100] crystal direction. (a) An $In_{0.53}Ga_{0.47}As$ configuration for a V_{GS} of 1.2 V and 1.7 V and (b) a Si configuration for a V_{GS} of 1.5 V and 2 V. Indicated schematically are estimations of the first three quantized energy levels for the highest V_{GS}, corresponding to different valley effective masses inside the potential well. The dashed line represents the quasi Fermi level for the holes in the source (E_{Fp}).

due to the strong FIQC, several distinct subband energy ladders can appear in the well, since the degree of confinement depends on the effective masses of the different band structure valleys. E.g. in a Si nTFET, those conduction band valleys which have a heavy longitudinal effective mass in the direction of confinement, are shifted in energy less than those for which the lower transverse mass determines the confinement. This leads to two ladders of quantized energy levels, as shown in Fig. 2.7(b). As a consequence, the tunneling currents to the different valleys see a relative shift, which is visible in the transfer characteristics as a kink in the SS [1]. This relative shift also affects the pTFET, in which the light hole band is shifted more by the confinement than the heavy hole band. Tunneling between conduction band and heavy hole band is not as efficient as to the light hole band: because of symmetry reasons, there is no direct coupling, so any tunneling transition must be assisted by a phonon. In the absence of confinement, this inefficient tunneling current is masked by the higher light hole-conduction band current. In the presence of FIQC, however, the relative shift results in the heavy hole-conduction band current forming an undesired tail to the transfer characteristics [42].

2.4.3 Point and line tunneling combined

Point and line tunneling current components can be present together in the same configuration. This is certainly true if the gate overlaps both the source and the channel region, but can also be induced by the fringing field of the gate, even if no gate-channel overlap is present. Because the line tunneling is more impacted by FIQC, V_{onset} for the point tunneling component is lower, meaning the gradual onset of the point tunneling can degrade the abrupt line tunneling onset. On the other hand, however, a gate-channel overlap removes the potential barrier between the source and the ungated channel, which can impede carriers from flowing to the drain and hence can decrease I_{ON} [43]. This introduces a trade-off between SS and I_{ON} and makes the alignment of the gate to the source-channel junction an important potential source for device variability in a lineTFET. However, the unwanted lateral tunneling component and the associated variability can be removed with the introduction of a counterdoped pocket, as discussed in Section 2.8.2.

2.5 Modeling

To acquire more physical insight into TFET operation and assess different architecture options, several models have been established. Since tunneling is a quantum mechanical phenomenon, all of these models rely on the solution

of some form of the Schrödinger equation. Even so, the different solution approaches can be categorized as either semiclassical (Section 2.5.1) or fully quantum mechanical (Section 2.5.2). In semiclassical models, the Schrödinger equation is not solved directly. Rather, the tunneling probability is calculated based on the integral of a position dependent imaginary wave vector along a well-defined tunnel path. This implies the classical assumption that wave vector and position of the electron are known at the same time, violating the Heisenberg uncertainty principle. In fully quantum mechanical approaches, on the other hand, the electron is described entirely by its wave function, which is obtained from a solution of the Schrödinger equation projected on a chosen basis. Tunneling then arises as a consequence of the wave-like character of the electron. Here, the most common examples of both approaches are discussed, without being exhaustive.

2.5.1 Semiclassical

A common semiclassical approach to the solution of the time-independent Schrödinger equation is the Wentzel-Kramers-Brillouin (WKB) approximation. The reasoning behind WKB starts from the one-electron wave function ψ in zero electric field, which corresponds to a constant potential energy. Assuming the electric field to be zero in the tunneling direction x, and disregarding the other directions for now, ψ takes the form of a plane wave [44]:

$$\psi(x) = A\exp\left(\pm ik_x x\right) \tag{2.5}$$

with i the imaginary number, x the tunneling direction, k_x the wave number and A the amplitude. The plus (minus) sign corresponds to a right (left) moving wave. The approximation then lies in assuming that the wave function in the presence of a small and smoothly varying non-zero field, can be described by introducing a position dependence for k_x. It can be shown that the phase $\phi(x)$ of the wave function can be obtained from the integral of $k_x(x)$ over the given domain [44]:

$$\psi(x) \equiv A(x)\exp(i\phi(x)) \approx \frac{C}{\sqrt{|k_x(x)|}}\exp\left(\pm i\int k_x(x)dx\right) \tag{2.6}$$

with C a real constant. In a forbidden energy region, like the bandgap, k_x is imaginary ($k_x = i\kappa_x$), which results in an exponential decay of the wave function. This corresponds to a tunneling process. By comparing the probability density at each side of the tunneling barrier, an expression for the transmission probability can be derived. This expression typically ignores the prefactors of the exponentials and therefore solely consists of a contour integral of the

imaginary k_x along the tunnel path through the forbidden region:

$$T_{\mathrm{WKB}} = \exp\left(2\int_{x_1}^{x_2} \kappa_x(x)dx\right) \tag{2.7}$$

with x_1 and x_2 respectively the start and end point of the tunnel path, also known as the classical turning points [44, 45, 46]. The tunnel path starts at the valence band edge and ends at the conduction band edge for a particular set of perpendicular wave numbers k_y and k_z. For zero perpendicular momentum ($k_y = k_z = 0$ and $\kappa_x = \kappa_{x0}$), the application of Eq. (2.7) is illustrated in Fig. 2.8. For non-zero k_y and k_z, the effective tunnel gap increases, making transmission less probable. This effect can be made explicit by rewriting Eq. (2.7) as:

$$T_{\mathrm{WKB}}(k_y, k_z) = \exp\left(2\int_{x_1}^{x_2} \kappa_{x0}(x)dx\right) \exp\left(-|k_y^2 + k_z^2| \int_{x_1}^{x_2} \frac{dx}{\kappa_{x0}(x)}\right) \tag{2.8}$$

where the assumption has been made that $k_y^2 + k_z^2 << k_{\mathrm{tot}}^2$ [46]. It is now clear that the second exponential factor of Eq. (2.8) reduces the transmission for non-zero perpendicular momentum [45, 46]. In the case of direct BTBT, κ_{x0} can be extracted from the complex band structure of the material under study (see Fig. 2.8(a) and Fig. 2.9 for example band structures). Although the WKB method can describe quantum phenomena like tunneling, it is still a semiclassical approach, because Eq. (2.8) requires that both position and momentum (expressed by the wave vector) are known at the same time. This is possible in a classical approach, but violates the quantum mechanical Heisenberg uncertainty principle. From the WKB transmission probability in Eq. (2.7), the BTBT current can be calculated. This is discussed later in this section.

Another commonly used semiclassical model to calculate the transmission probability is Kane's model [47]. Originally, the Kane formula for BTBT probability was derived for a uniform electric field in a perturbative approach, using Fermi's golden rule and assuming a 2-band $\mathbf{k}\cdot\mathbf{p}$-model. An equivalent result can be obtained by starting from the WKB method and assuming the electric field F to be constant over the tunnel path length, while taking a two band $\mathbf{k}\cdot\mathbf{p}$ model to describe the complex wave vector dispersion. The transmission probability in a direct bandgap material can then be written as:

$$T_{\mathrm{Kane}}(k_y, k_z) = \exp\left(-\frac{\pi E_{\mathrm{G}}^{3/2} m_{\mathrm{R}}^{1/2}}{2q\hbar F}\right) \exp\left(-2\frac{E_\perp}{\overline{E}}\right) \tag{2.9}$$

where E_{G} is the effective bandgap at the tunnel junction and m_{R} is the reduced effective mass, defined as $m_{\mathrm{R}} = \frac{m_{\mathrm{el}} m_{\mathrm{lh}}}{m_{\mathrm{el}} + m_{\mathrm{lh}}}$, with m_{el} and m_{lh} the effective masses of respectively the conduction band and the light hole band. In the second exponential factor, $E_\perp = \hbar^2(k_y^2 + k_z^2)/(2m_{\mathrm{R}})$ and $\overline{\mathrm{E}} = 2q\hbar F / \left(\pi (m_{\mathrm{R}})^{1/2} E_{\mathrm{G}}^{1/2}\right)$.

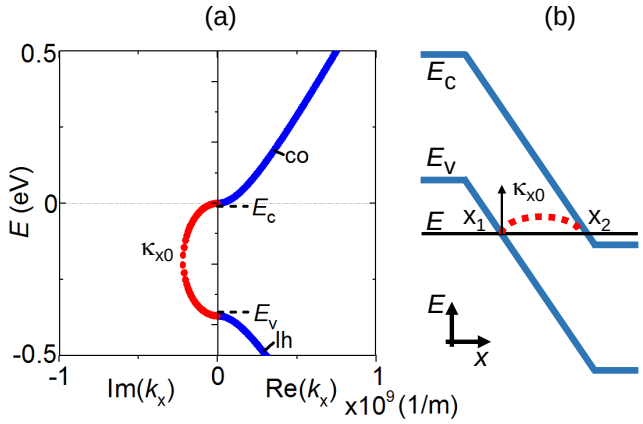

Figure 2.8: (a) Real and imaginary two-band **k**·**p** band structure of bulk InAs for zero perpendicular momentum ($k_y = k_z = 0$). Indicated are the conduction (co) and light hole (lh) bands. (b) Energy band diagram, superimposed with the imaginary band structure at a given energy E. The WKB transmission probability is calculated from an integral of this imaginary dispersion.

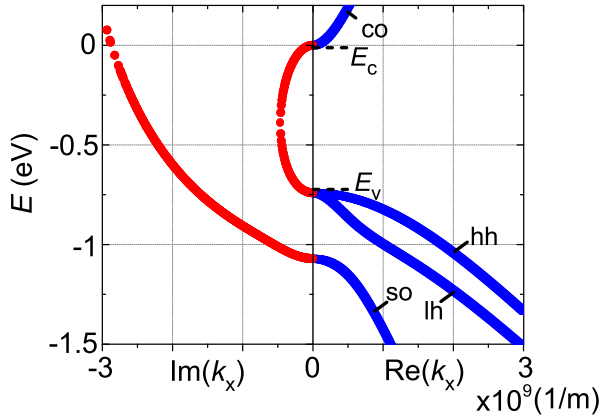

Figure 2.9: Real and imaginary thirty-band **k**·**p** band structure of bulk $In_{0.53}Ga_{0.47}As$. Indicated are the conduction (co), light hole (lh), heavy hole (hh) and split-off (so) bands.

A factor $\frac{\pi^2}{9}$, present in the original Kane derivation, was shown later to be incorrect and has been removed [48]. Because of the assumption of a constant electric field, the Kane model can deviate significantly from the WKB approach in cases where the field is strongly non-uniform [48, 49].

From the transmission probabilities in Eq. (2.7) and Eq. (2.9), the BTBT current can be calculated in two ways. The first is based on the ballistic Landauer formalism and entails a direct integration of the transmission probabilities, weighted with the distribution functions in the contacts. This is called the Tsu-Esaki formula [50], which gives the BTBT current density as:

$$I_{\mathrm{DS}} = \frac{2q}{h} \int_{k_x,k_y,k_z} T(k_y,k_z)(f_{\mathrm{S}}(E) - f_{\mathrm{D}}(E)) \frac{dk_x}{2\pi} \frac{dk_y}{2\pi} \frac{dk_z}{2\pi} \tag{2.10}$$

where T is the transmission probability and $f_{\mathrm{S}}(E)$ and $f_{\mathrm{D}}(E)$ are the Fermi-Dirac distributions in the source and drain contacts respectively. The energy E is a function of k_x, k_y and k_z. Eq. (2.10) illustrates that the BTBT current is determined by both the transmission probability and the occupation probabilities at both sides of the tunnel junction.

A second way to calculate the BTBT current, with stronger simplifications, is to determine a generation rate per unit volume, which is then integrated over the full device volume. In this approach, device simulators search for tunnel paths which connect points of sufficient potential difference to allow for tunneling. Carriers are generated by the BTBT process at the endpoints of the tunnel paths. An expression for the generation rate can be derived from Eq. (2.10) by assuming the distribution functions for the carriers to be step functions, corresponding to a temperature of 0 K. This effectively decouples the carrier distribution functions from the transition rates. For the Kane model, the following generation rate formula is commonly used:

$$G_{\mathrm{Kane}} = A \left(\frac{F}{F_0}\right)^D \exp\left(\frac{-B}{F}\right) \tag{2.11}$$

where F is the electric field, F_0 is 1 V/cm, D is a parameter that is taken 2 for direct bandgap materials and A and B are parameters defined as:

$$A_{\mathrm{direct}} = \frac{g m_{\mathrm{R}}^{1/2} (qF_0)^2}{\pi h^2 (E_{\mathrm{G}})^{1/2}} \tag{2.12}$$

$$B_{\mathrm{direct}} = \frac{\pi^2 m_{\mathrm{R}}^{1/2} (E_{\mathrm{G}})^{3/2}}{qh} \tag{2.13}$$

where g is a factor for the spin and valley degeneracies. Corrections that reintroduce the non-zero temperature distribution functions afterwards have

been developed [51]. Extensions of Kane's model have been made to describe indirect phonon-assisted transitions as well [45]. For indirect BTBT, D in Eq. (2.11) is 2.5 and A and B are modified to [43]:

$$A_{\text{indirect}} = \frac{g(m_{\text{c,DOS}}m_{\text{v,DOS}})^{3/2}(1+2N_{\text{PH}})D_{\text{PH}}^2(qF_0)^{5/2}}{2^{21/4}h^{5/2}m_{\text{R}}^{5/4}\rho\epsilon_{\text{PH}}E_{\text{G}}^{7/4}} \tag{2.14}$$

$$B_{\text{indirect}} = \frac{2^{7/2}\pi m_{\text{R}}^{1/2}E_{\text{G}}^{3/2}}{3qh} \tag{2.15}$$

where $m_{\text{v,DOS}}$ ($m_{\text{c,DOS}}$) is the valence (conduction) band density of states effective mass, $N_{\text{PH}} = 1/\left[\exp(\epsilon_{\text{PH}}/kT) - 1\right]$, D_{PH} and ϵ_{PH} are respectively the occupation number, the deformation potential and the energy of the relevant phonons and ρ is the mass density. To obtain the BTBT current, the generation rate of Eq. (2.11) is integrated over the device volume:

$$I_{\text{DS}} = q\int GdV \tag{2.16}$$

with dV an elementary volume. The same approach can be followed for the WKB approximation [51].

The disadvantage of semiclassical methods is that they neglect certain quantum phenomena which result from the wave-like character of the electron, since they do not directly solve the Schrödinger equation in the full device region. Important examples for TFET include field or size-induced confinement effects, which can effectively increase the bandgap, and resonances and reflections in regions of high field, which respectively increase or decrease the transmission probability [52, 53].

2.5.2 Quantum mechanical

A fully quantum mechanical simulation approach entails the solution of the Schrödinger equation in some form. In TFET modeling, the most commonly used is the time independent one-electron form:

$$\hat{H}\psi(\mathbf{r}) = \left[\frac{-\hbar^2}{2m_e}\nabla^2 + V_{\text{e}}(\mathbf{r}) + V_{\text{c}}(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{2.17}$$

with m_e the free electron mass, E the total energy, $V_{\text{e}}(\mathbf{r})$ the external applied potential energy and $V_{\text{c}}(\mathbf{r})$ the crystal potential energy of the lattice. The solution of Eq. (3.1) generally occurs in two steps. First, the wave function is decomposed onto a reduced basis. This is because the complicated nature of $V_{\text{c}}(\mathbf{r})$

prevents the direct solution of Eq. (3.1) for ψ. The choice of basis corresponds to a particular band structure model. For TFET, the most commonly used models are the **k**·**p**-based envelope function (EF) method and the orbital-based tight-binding (TB) method. Once the Hamiltonian has been written in the chosen basis, the second step is to construct a linear system or an eigenvalue problem, the solutions of which can be used to extract desired quantities like currents and carrier densities. The two main approaches for this step are the wave function (WF) method and the non-equilibrium Green's function (NEGF) method.

A first band structure model is the EF method, which expands the wave function on the solutions of the Schrödinger equation in bulk [54]:

$$\psi(\mathbf{r}) = \sum_n F_n(\mathbf{r}) U_n(\mathbf{r}), \tag{2.18}$$

where the $U_n(\mathbf{r})$ form a complete set of orthonormal basis functions with the periodicity of the lattice and $F_n(\mathbf{r})$ are slowly varying envelope functions, which contain only Fourier components in the first Brillouin zone. The index n runs over all bands considered in the description. Inserting the expansion of Eq. (3.4) into Eq. (3.1) and using the properties of the set $U_n(\mathbf{r})$, namely orthonormality and completeness, the following system of equations results (a detailed derivation can be found in the work of Burt [54] and Van de Put *et al.* [52]):

$$\frac{-\hbar^2}{2m_e}\nabla^2 F_n(\mathbf{r}) - \frac{i\hbar}{m_e}\sum_m \mathbf{p}_{nm}\cdot\nabla F_m(\mathbf{r}) + \sum_m H_{nm}(\mathbf{r})F_m(\mathbf{r}) + V_{\mathrm{e}}(\mathbf{r})F_n(\mathbf{r}) = EF_n(\mathbf{r}) \tag{2.19}$$

where the external potential V_{e} is assumed to vary slowly on the scale of a unit cell, like the EFs. The $\mathbf{p}_{nm}$ are known bulk **k**·**p** interband momentum matrix elements. They describe the coupling strength, and hence BTBT, between bands n and m. H_{nm} are the bulk Hamiltonian matrix elements, which correspond to known bulk band edge energies. The main advantage of Eq. (2.19), compared to Eq. (3.1) is therefore that the crystal potential $V_{\mathrm{c}}(\mathbf{r})$ has been replaced with known material parameters.

An alternative band structure description, called the TB method, consists of expanding the wave function on Bloch sums of localized atomic orbital-like functions, instead of on extended bulk solutions [19, 55, 56]:

$$\begin{aligned}\psi(\mathbf{r}) &= N^{-\frac{1}{2}}\sum_n C_n \sum_i \exp(i\mathbf{k}\cdot\mathbf{R}_i)\phi_n(r-\mathbf{R}_i)\\ &= N^{-\frac{1}{2}}\sum_{n,i} C_{n,i}\phi_n(r-\mathbf{R}_i)\end{aligned} \tag{2.20}$$

where ϕ_n is a Löwdin orbital with quantum number n, located on the atom at position $\mathbf{R}_i$. N is the number of primitive unit cells in the crystal and serves as a normalization constant. $C_{n,i} = C_n \exp(i\mathbf{k} \cdot \mathbf{R}_i)$ are the expansion coefficients, which serve a similar purpose as the EFs F_n of Eq. (3.4). Inserting the expansion of Eq. (2.20) into the Schrödinger equation Eq. (3.1), multiplying from the left with the Bloch sum $\sum_j \exp(-i\mathbf{k} \cdot \mathbf{R}_j)\phi_m^*(\mathbf{r} - \mathbf{R}_j)$ and integrating over the full crystal, the following system of equations is obtained:

$$\sum_{\mathrm{i,j}} C_{n,i-j} \int \phi_m^*(r - \mathbf{R}_j)\hat{H}\phi_n(r - \mathbf{R}_i)d\mathbf{r} = \sum_{i,j} C_{n,i-j}E \tag{2.21}$$

with $C_{n,i-j} = C_n \exp(i\mathbf{k}\cdot(\mathbf{R}_j - \mathbf{R}_i))$. The integrals in Eq. (2.21) can be replaced by parameters available in literature, similar to the $\mathbf{p}_{nm}$ elements in the EF approach.

For both the EF method and the TB method, the inclusion of more basis functions results in a more accurate band structure that can capture a larger part of the first Brillouin zone. If enough bands are included such that the full first Brillouin zone is captured, the model is called a full-zone model. For the EF method, this corresponds to a 30-band basis [57, 58], for the TB method to the $\mathrm{sp^3d^5s^*}$ basis set [59]. With more basis functions, also the coupling between the different bands is captured more accurately (compare e.g. Fig. 2.8 with Fig. 2.9).

To solve Eq. (2.19) or Eq. (2.21) for an actual device in which current enters and leaves through the source and drain contacts, open boundary conditions are required. For TFET simulations, the quantum transmitting boundary method (QTBM) is often used [60]. QTBM imposes that the electric field is zero in the contacts in the transport direction. The wave function is then known to take the form of a plane wave in that direction. It is this form which is imposed as a boundary condition on the system.

With the appropriate boundary conditions, Eqs. (2.19) and (2.21) can be discretized and solved for the EFs and TB coefficients respectively. This direct solution approach is called the WF method. Available numerical techniques include finite difference (FD) methods [61], finite element (FE) methods [62], spectral methods [63] and others. The resulting EFs and TB coefficients can be used to calculate important device characteristics such as carrier densities, transmission probabilities and currents. Similar to the semiclassical case, the current can be calculated from the transmission probabilities directly with Eq. (2.10), or more approximative by first converting to a generation rate and using Eq. (2.16). To include the effect of the carrier concentration on the potential energy profile, the calculated carrier densities can be used to calculate a new potential energy profile with the Poisson equation, which in its turn can

be used to calculate a new carrier density. This loop is then repeated until self-consistency is reached.

Instead of solving Eqs. (2.19) and (2.21) directly as in the WF method, an alternative is to calculate the non-equilibrium Green's function (NEGF) [56, 64]. In an NEGF approach, Eq. (3.1) is rewritten as:

$$(EI - \hat{H} - \hat{\Sigma}_{\mathrm{S/D}})G = I \tag{2.22}$$

with G the Green's function, $\hat{H}$ the Hamiltonian of the closed system in the chosen basis and $\hat{\Sigma}_{\mathrm{S/D}}$ the self-energies of the source and drain contacts. The self-energy terms are added to include the open boundary conditions. Just like the wave function, G contains all relevant device information. The advantage of the NEGF approach is that other interactions can be included in a rather straightforward manner by adding the corresponding self-energy term. E.g. for electron-phonon interactions:

$$(EI - \hat{H} - \hat{\Sigma}_{\mathrm{S/D}} - \hat{\Sigma}_{\mathrm{el-ph}})G = I \tag{2.23}$$

with $\hat{\Sigma}_{\mathrm{el-ph}}$ the self-energy for the electron-phonon interaction [56]. In this way, rethermalization, phonon-assisted tunneling and other scattering processes can be described. To avoid having to calculate the inverse of a large matrix $(EI - \hat{H} - \hat{\Sigma}_{\mathrm{S/D}} - \hat{\Sigma}_{\mathrm{el-ph}})$, methods have been developed that calculate only the relevant entries in the Green's function matrix, such as the contact block reduction (CBR) method [65] and the recursive Green's function (RGF) method [56].

A separate quantum-mechanical framework based on wave functions has also been established for phonon-assisted tunneling [1]. It was shown that the current can be determined from solving Eq. (2.19) for each band separately. Two distinct sets of wave functions are then obtained for the conduction band and the valence band. The transmission probability is calculated based on the overlap of the wave functions of the two bands, combined with the interaction probability with a phonon of the appropriate momentum.

2.5.3 Semiclassical versus quantum mechanical

Compared to semiclassical approaches, quantum mechanical models are typically computationally more expensive, but include the effects of the wave-like nature of the electron, such as confinement and reflections in regions of high field or at a heterojunction. Quantum simulations should therefore be preferred in cases where confinement is expected to be strong, e.g. for TFETs with a T_{body} below the Bohr radius of the material, or when the tunneling is oriented toward the

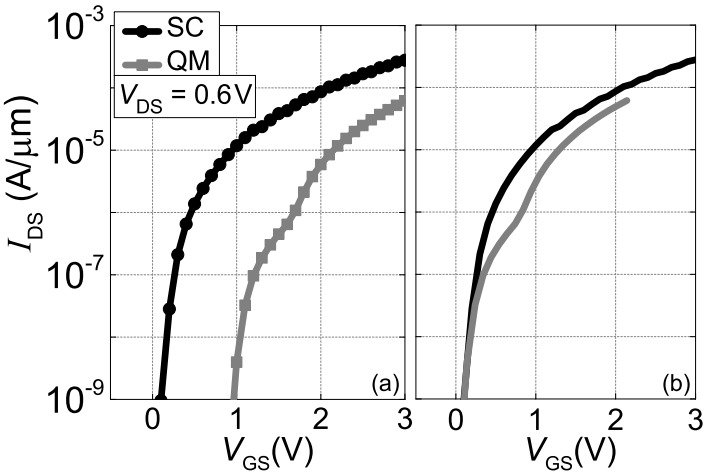

Figure 2.10: Transfer characteristics of a Si lineTFET simulated with a semiclassical (SC) approach based on the WKB method and a quantum mechanical (QM) approach based on wave function overlap. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is 1x10^{-9}A/μm coincides.

quantum well underneath the gate dielectric. Other cases include TFETs with a heterojunction [18] or configurations that rely on resonant tunneling [66].

To illustrate the differences that can exist between semiclassical and quantum mechanical approaches, Fig. 2.10 compares the simulated transfer characteristics of a Si lineTFET, whereby the last described quantum mechanical approach for phonon-assisted tunneling is used. Since the tunneling is oriented towards the gate, a strong impact of FIQC is expected. The effect of FIQC is visible for the quantum simulation as a shift in V_{onset} compared to the semiclassical prediction and a kink due to different subband ladders (see also Section 2.4.2).

2.6 Challenges

Having a basic understanding of TFET and methods to model the tunneling current, the main challenges for TFET implementation are now discussed in more detail. Although the operating principle of the TFET is promising

for low power applications, several important challenges remain. They can be summarized as reaching a high I_{ON} (Section 2.6.1), combined with a low SS over several decades of current (Section 2.6.2), while maintaining a low I_{OFF} (Section 2.6.3). These three device characteristics are discussed with the appropriate metrics and the parameters which have an influence on them.

2.6.1 I_{ON}

The TFET I_{ON} is typically lower than that of a MOSFET. This is due to the tunneling barrier, which all charge carriers have to overcome before they can drift and diffuse toward the drain. This barrier is present even in the ON-state of the device. In contrast, in a MOSFET, the carriers with an energy higher than the source-channel barrier can drift and diffuse from source to drain unimpeded. The lack of sufficient drive current negatively impacts the intrinsic delay of a TFET inverter configuration, expressed as:

$$\tau_{\mathrm{d}} = \frac{C_{\mathrm{ox}} V_{\mathrm{DD}}}{I_{\mathrm{ON}}} \tag{2.24}$$

From Eq. (2.24), it is clear that to keep the delay small in low-V_{DD} operation, I_{ON} should be sufficiently high. In order to compete with MOSFET, it is generally accepted that I_{ON} should be in the range of several hundreds of μA/μm for V_{DD} smaller than 0.5 V [23].

To identify the parameters that increase I_{ON}, the Kane model can be used, although it is strictly speaking only valid for uniform fields. The exponential factors in Eqs. (2.11)-(2.15) show that the BTBT generation rate, and hence I_{ON}, is largest for a high electric field at the tunnel junction, a small bandgap and a small reduced effective mass. A high F and small E_{G} correspond physically to short tunnel paths, while the low m_{R} signifies a small attenuation along the path, as it is correlated with the imaginary dispersion in the bandgap. A short tunnel path and a small attenuation result in a high probability for the charge carriers to tunnel into the channel and contribute to the current. A high F can be obtained with sharp doping profiles and a large doping level at the tunnel junction, by increasing the source doping and/or with the introduction of a counterdoped pocket (see Section 2.8). Additionally, F can be improved by enhancing control of the gate over the tunnel junction, e.g. by thinning the device body, by adding gates to obtain a multi-gate (MuG) configuration or by modifying the device structure to a line tunneling set-up (see Sections 2.4 and 2.9). The small E_{G} and reduced m_{R}, on the other hand, are the result of a proper material choice, with III-V materials proving promising candidates, either in a homostructure (one material) or in a heterostructure (multiple materials) configuration (see Section 2.7).

2.6.2 SS

In addition to a high I_{ON}, a TFET should obtain a low, sub-60 mV/dec SS over a large current range of interest. The SS of a TFET at a specific V_{GS} can be approximated as:

$$SS = \frac{dV_{GS}}{d\log_{10}(I_{DS})} = \ln(10)\left[\frac{1}{V_R}\frac{dV_R}{dV_{GS}} + \frac{F+B}{F^2}\frac{dF}{dV_{GS}}\right]^{-1} \tag{2.25}$$

which is derived from an approximation of the tunnel junction as a degenerately doped p-n junction with a constant electric field [67]. V_R is then the reverse bias of the tunnel junction (E_{Fp}-E_{Fn}) and B is the exponential parameter of the Kane formalism (see Eq. (2.13) and Eq. (2.15)). Comparing Eq. (2.25) to Eq. (2.4) shows that in contrast to the MOSFET, the SS of a TFET is indeed not limited by $\frac{kT}{q}\ln(10)$. However, Eq. (2.25) also shows that the TFET SS is not constant with V_{GS}. It is therefore possible that the sub-60 mV/dec SS regime is only limited to very small levels of I_{DS}. In this case, reducing the supply voltage is still not feasible, as the majority of the SS is above the thermionic limit. The minimum value of the SS as described in Eq. (2.25) at a particular V_{GS}-point, also called the minimum point-SS, is therefore not a good metric to assess TFET performance. More relevant is to look at the average SS over the full operating voltage range, as shown in Fig. 2.11:

$$SS_{avg} = \frac{V_{DD} - V_{OFF}}{\log_{10}\left(\frac{I_{ON}}{I_{OFF}}\right)} \tag{2.26}$$

where V_{OFF} is the V_{GS} corresponding to the defined I_{OFF}. SS_{avg} still depends, however, on the choice of I_{OFF} and V_{DD}, which can vary depending on the targeted application. Therefore, another metric called I_{60} has been proposed [68]. I_{60} is defined as the point on the I_{DS}-V_{GS} curve where the SS transitions from sub-60 mV/dec to super-60 mV/dec (also indicated in Fig. 2.11). Hence, it is the largest current for which the corresponding TFET outperforms a theoretically optimal MOSFET. I_{60} is independent of the choice of I_{OFF}, V_{DD} or the gate work function, so it can be used to compare various TFET configurations presented in literature. To be competitive with MOSFET, I_{60} should be at least above 10 μA/μm [23, 68]. As an example, Table 2.1 lists some experimental values obtained in literature. See also Section 2.12 for more experimental results. These examples show that there is still a large gap with the SS goal.

From Eqs. (2.25) and (2.26), it is clear that SS can be minimized by increasing $\frac{dV_R}{dV_{GS}}$ and $\frac{dF}{dV_{GS}}$, which represent the gate control over the tunnel junction, and by increasing the ratio I_{ON}/I_{OFF}. Physically, these requirements correspond to the capability of the gate to induce an abrupt transition from an OFF-state

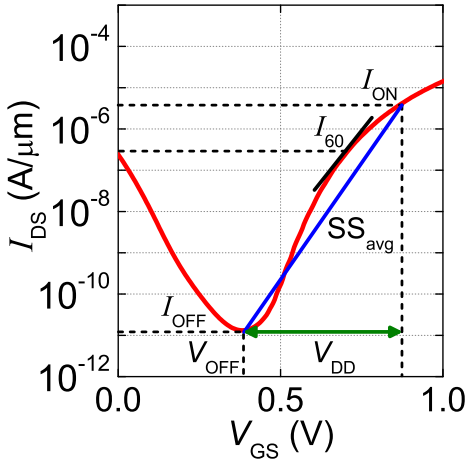

Figure 2.11: SC simulated transfer characteristics of an $In_{0.53}Ga_{0.47}As$ p-i-n TFET as shown in Fig. 2.3, illustrating various metrics to characterize TFET performance.

Table 2.1: Examples of recent experimental values for I_{60} in literature, rounded to order of magnitude. Target value for I_{60} is 10 μA/μm or larger.

nTFET	I_{60} [μA/μm]	pTFET	I_{60} [μA/μm]
Tomioka *et al.* [69]	10^{-3}	Mayer *et al.* [70]	10^{-6}
Dewey *et al.* [9]	10^{-3}	Morita *et al.* [71]	10^{-5}
Sarkar *et al.* [72]	10^{-4}		
Huang *et al.* [73]	$<10^{-6}$		
Kim *et al.* [74]	$<10^{-6}$		

with a low tunneling probability to an ON-state with a very high tunneling probability. This means the optimization of SS and I_{ON} are intertwined, with one generally benefiting from the improvement of the other. This is not true, however, for very high source doping levels. A large source doping improves I_{ON}, but can degrade the SS if the source degeneracy becomes too high, as was mentioned in Section 2.3. For a given doping level, the source degeneracy is determined by the density of states (DOS) in the valence band (for an nTFET) or the conduction band (for a pTFET) of the source material. A large DOS allows for a high source doping while keeping the source degeneracy, and hence the SS degradation, small. As the conduction band for most common bulk materials has a lower DOS than the valence band, the pTFET has a more limiting trade-off between I_{ON} and SS. This puts the pTFET at an inherent

disadvantage compared to the nTFET.

The SS can be negatively impacted by parasitic current mechanisms which occur in addition to the desired BTBT process, such as trap-assisted tunneling (TAT) [7, 75, 76]. TAT is a process in which trap levels in the bandgap are used by the charge carriers as stepping stones between conduction and valence band (see Fig. 2.12(a)). These levels are localized energy states resulting from defects in the material, such as vacancies, impurities and dangling bonds in the bulk or at interfaces. The TAT transition itself consists of at least one ballistic tunneling step and a phonon-assisted thermal step to and from the trap level. This means TAT has both a field and a temperature dependence. TAT degrades the SS, as it can turn on at lower V_{GS} and add a thermal tail to the transfer characteristics. The impact of TAT can be reduced by improving the material quality, such that the trap density is low, or by enhancing the desired BTBT current, such that it better masks the current component caused by TAT. The amount of TAT present in the transfer characteristics can be characterized by $V_{\mathrm{TAT}}@I_{\mathrm{OFF}}$, the increase in V_{onset} at 300 K at the specified I_{OFF} compared to the V_{onset} value at 77 K [77].

2.6.3 I_{OFF}

To obtain a low leakage power, I_{OFF} should be sufficiently small. I_{OFF} is composed of the reverse leakage current of a p-i-n diode, consisting mainly of SRH processes, minority carrier diffusion, and possibly some TAT if the other two mechanisms are sufficiently small. Additionally, parasitic TFET mechanisms such as ambipolar current and direct source-drain tunneling can increase I_{OFF}. For low-power applications, the ITRS roadmap sets the target for I_{OFF} at 10 pA/μm, with $\frac{I_{\mathrm{ON}}}{I_{\mathrm{OFF}}} > 10^5$.

SRH, similar to TAT, is a trap-assisted process in which carriers are generated via localized trap levels through phonon-assisted thermal steps when the electron and hole densities are out of equilibrium (see Fig. 2.12(b)). In contrast to TAT, SRH has a weak field dependence and so does not impact the SS. However, if the trap density is sufficiently high, it can increase I_{OFF}. Similar to TAT, the remedy is to decrease the trap density.

The ambipolar current is determined by the efficiency of BTBT at the channel-drain junction. This means it can be reduced by taking opposite measures as those suggested in Section 2.6.1 to increase I_{ON}, viz. a reduction of F and E_{G} and an increase in m_{R} at the channel-drain tunnel junction. F can be reduced by lowering the drain doping, at the expense of increasing the resistance. E_{G} and m_{R} depend on the material choice and cannot be varied independently from the source in a homostructure, but they can be in a heterostructure (see

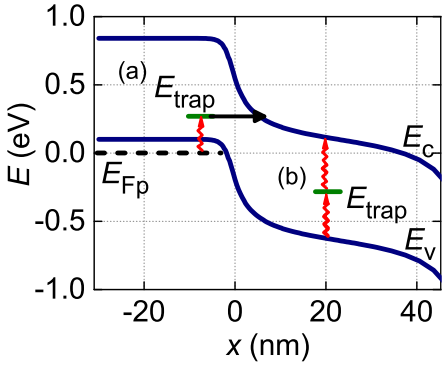

Figure 2.12: Energy band diagrams of the tunnel junction of an $In_{0.53}Ga_{0.47}As$ TFET showing (a) TAT and (b) SRH processes, which contribute to the TFET I_{OFF}. The short lines represent localized trap energy levels (E_{trap}), the wavy lines are thermal steps and the straight arrows correspond to a tunneling step.

Section 2.7.3). Another option is to introduce a gate-drain underlap region, which decreases the influence of the gate on the channel-drain junction [78] (see Section 2.9.2). Using a combination of these techniques, the ambipolar current can be kept to acceptable levels.

Direct source-drain tunneling occurs when charge carriers can tunnel from source to drain, even when the device is in the OFF-state. This occurs when the tunnel paths in the OFF-state are too short or provide insufficient attenuation to suppress the carrier tunneling probability. Direct tunneling is particularly important in materials with a small E_G and m_R and in configurations with a short channel. In vertical architectures, increasing the channel length is a straightforward way to reduce the source-drain tunneling. In planar configurations, however, a longer channel length also increases the device footprint.

2.7 Material choice

To overcome the challenges outlined in the previous section, different material options are being considered for TFET. Of particular interest are materials from group IV (Section 2.7.1) and compounds from group III and V of the periodic system of elements (Section 2.7.2). Multiple of these materials can be combined in a heterostructure (Section 2.7.3). 2D materials are a new class of materials, which is also gaining interest for TFET applications (Section 2.7.4).

2.7.1 Group IV materials

Silicon is the most prevalent material in today's MOSFETs for logic applications, and has therefore also been extensively studied for TFET. Si has the distinct advantage of being a well-known, abundant material that can be obtained with a very low density of defects in the bulk of the material as well as at the gate dielectric interface. Established fabrication infrastructure and processes exist in industry for a wide variety of process steps. This includes high quality oxide growth and high concentration doping. Si is also interesting for TFET because of its large conduction band DOS, which limits the source doping degeneracy in pTFETs (the importance of which is discussed in Section 2.6.2).

However, Si is a relatively poor material for BTBT and consequently, Si TFETs typically have a small I_{ON}. A first cause is the relatively large E_G of 1.12 eV, which results in a low tunneling probability (see Section 2.6.1). Although beneficial for I_{OFF}, the low tunneling rate is detrimental for I_{ON}. Secondly, Si is an indirect bandgap material, which reduces the overall efficiency of BTBT. The smallest E_G is measured between the valence band maximum at the Γ-point and the conduction band valley minimum along the Δ-direction [57]. These two points in the E-k diagram are not at the same k-value, which means they correspond to a different crystal momentum $\hbar k$. A tunneling transition between these two points requires a phonon to provide the necessary crystal momentum. This makes it less probable for a given energetic separation than a direct transition, in which the start and end points of the tunneling process are at the same k-value, therefore requiring no phonon. In Si, the indirect transitions are dominant, since the bandgap at the Γ-point is much larger at 3.4 eV, making direct transitions very improbable [43].

As an alternative to Si in group IV of the periodic system of elements, Ge is more suited for BTBT. It has a smaller indirect bandgap than Si of 0.66 eV. Unlike Si, the direct bandgap at the Γ-point is only slightly larger at 0.8 eV. Consequently, once the band bending is such that direct transitions are allowed, they will dominate over the indirect transitions and improve I_{ON} [79, 43]. However, as discussed in Section 2.6.3, a smaller bandgap facilitates ambipolar tunneling at the channel-drain junction, which is an important leakage mechanism in TFET. I_{OFF} will therefore generally be larger than for Si implementations, but it is expected that acceptable levels of I_{OFF} can be reached.

Since it belongs to the same group as Si, Ge can also be used in an alloy with Si, forming Si_xGe_{1-x}, with x the Si mole fraction. x determines the band structure, including the bandgap and the relative importance of the indirect and direct processes. A rather abrupt transition occurs from Ge-like to Si-like behavior above x=0.2 [43]. The control over the band structure through x can be used

to find a compromise between improving I_{ON}, and reducing I_{OFF}.

2.7.2 Group III-V materials

Beyond group IV, a host of materials suited for TFET can be found in the class of III-V materials, which are compounds of elements from group III and group V. They are actively researched as future channel materials for MOSFET, because they can exhibit high electron mobilities compared to Si [80]. III-V materials show a wide variety of band structures, with bandgaps varying from 0.17 eV for InSb up to 3.28 eV for GaN [81]. The same variety can be found in the effective masses. The possibility of a small bandgap and small effective mass, combined with the direct nature of most III-V materials, makes them interesting candidates for improving the TFET I_{ON}. Important examples which have been studied for TFET applications are binary compounds like InAs, GaSb, InP and ternary compounds like $In_xGa_{1-x}As$ and $GaAs_xSb_{1-x}$ [26].

Nevertheless, III-V materials also present significant challenges. Firstly, they are much less known than Si, so it is challenging to obtain high-quality materials and gate dielectrics with a low defect density [80]. Secondly, the material properties that allow for a high I_{ON} also raise I_{OFF} undesirably. A small bandgap combined with a low effective mass facilitates both ambipolar and direct source-drain tunneling. This issue can be alleviated with the use of a heterostructure as discussed in Section 2.7.3. Thirdly, a low electron effective mass means a low DOS in the conduction band compared to Si. This exacerbates the I_{ON}-SS trade-off for pTFET for increasing source doping, discussed in Section 2.6.2.

2.7.3 Heterostructures

The material options discussed so far assume that the active region of the device consists of only one material. For such a configuration, an important issue in the material choice is the link between I_{ON} and I_{OFF}, with the desired improvement of the former also inducing an unwanted increase in the latter. A way to decouple the two is to combine two materials in a heterostructure, such that the material in the source is different from the material in the channel and the drain.

A heterostructure configuration has an extra degree of freedom: the band edge alignment at the tunnel junction. The alignment is determined to a first approximation by the electron affinity of the two constituent materials, and is classified as straddled, staggered or broken, as shown in Fig. 6.15. The effective bandgap which determines the tunnel path lengths at the tunnel junction

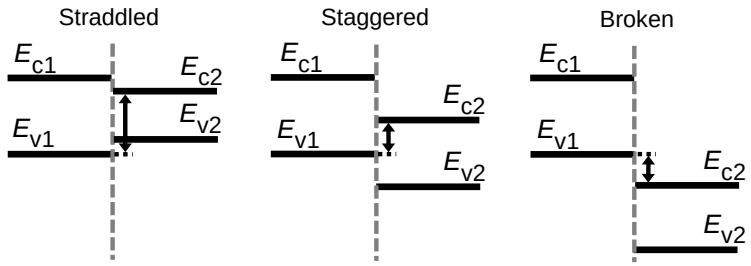

Figure 2.13: Types of band edge alignments at a heterostructure tunnel junction. In a straddled alignment, $E_{c1} > E_{c2}$ and $E_{v1} < E_{v2}$. In a staggered alignment, $E_{c1} > E_{c2}$ and $E_{v1} > E_{v2}$. In a broken alignment, $E_{v1} > E_{c2}$. The double arrows indicate the effective bandgap at the tunnel junction. For the broken configuration, the effective bandgap is negative.

is determined by the alignment and not by the bandgaps of the individual materials. A hetero-TFET can therefore combine a very small effective bandgap at the source-channel junction, enabling a high I_{ON}, with a large bandgap at the channel-drain junction, maintaining a low I_{OFF}. Fig. 2.14 illustrates the performance improvement for a staggered heterojunction compared to a homojunction. The material of the source can also be chosen to have a large DOS, associated with a large bandgap, to limit the source doping degeneracy. This can be done without increasing the effective bandgap at the tunnel junction, so long as the band alignment remains favorable [5].

In the special case of a broken band alignment, there is no forbidden region through which the carrier has to tunnel. This enables a high transmission probability and is hence beneficial for I_{ON}. It is still unclear, however, whether a broken gap alignment will allow for a sufficiently low I_{OFF}, since phonon-assisted leakage paths exist in the OFF-state [82, 83]. This is indicated in Fig. 2.15: after transitioning from source to channel, a rethermalization step can enable the electron to pass over the potential barrier in the channel, resulting in an increased I_{OFF}. The importance of this leakage current is influenced by the quantization in the triangular wells at the junction and the interaction strength with the available phonons.

Depending on the choice of materials, heterostructures can be either lattice-matched or lattice-mismatched. In the first case, both materials have the same lattice constant. The materials can be grown epitaxially without any stress at the heterojunction. Notable examples include $In_{0.53}Ga_{0.47}As/InP$ (straddled) and $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ (staggered). In the second case, the mismatch of lattice constants results in a non-uniform stress profile arond the tunnel

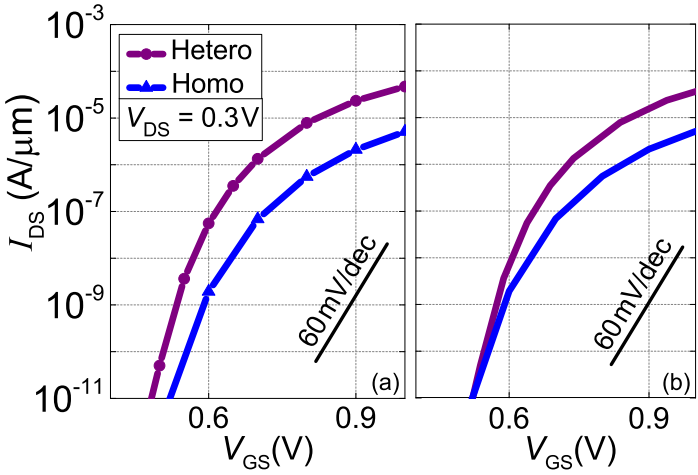

Figure 2.14: QM simulated transfer characteristics for a $In_{0.53}Ga_{0.47}As$ (bandgap of 0.74 eV) homostructure TFET and a $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ heterostructure TFET. The heterostructure is lattice-matched and has a staggered band alignment with an effective bandgap of 0.29 eV. The TFET is a double-gate configuration with an EOT of 0.6 nm, a T_{body} of 10 nm and a source doping of $5x10^{19}cm^{-3}$. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1x10^{-11}A/\mu m$ coincides. This shift can be accomplished by a proper choice of gate work function.

junction, as in the case of GaSb/InAs (broken). The stress has an impact on the local band structure (see Section 2.10). The build-up of stress can also result in interfacial defects at the junction if relaxation sets in [84].

2.7.4 2D materials

2D materials are an alternative to the bulk semiconductors described in the previous sections, and are being investigated for both MOSFET and TFET applications [89]. 2D materials are atomically thin, enabling excellent gate control. In addition, it is expected that low defect densities can be obtained with no dangling bonds or roughness at the surface.

The most studied 2D-material is graphene, but since it is a semi-metal, it does not have a bandgap and is therefore not suited for TFET as such. However, a

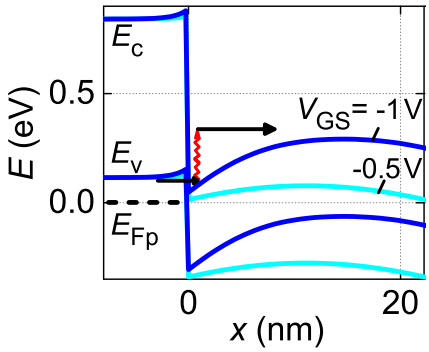

Figure 2.15: Energy band diagram at the tunnel junction of a GaSb/InAs heterostructure TFET with a broken band alignment in the OFF (dark lines) and ON-state (light lines). V_{GS} is -1 V and -0.5 V respectively. Indicated is a leakage path in the OFF-state. The wavy line represents a thermal step. The dashed line represents the quasi Fermi level for the holes in the source (E_{Fp}). The source doping is 5x10^{19}cm^{-3}.

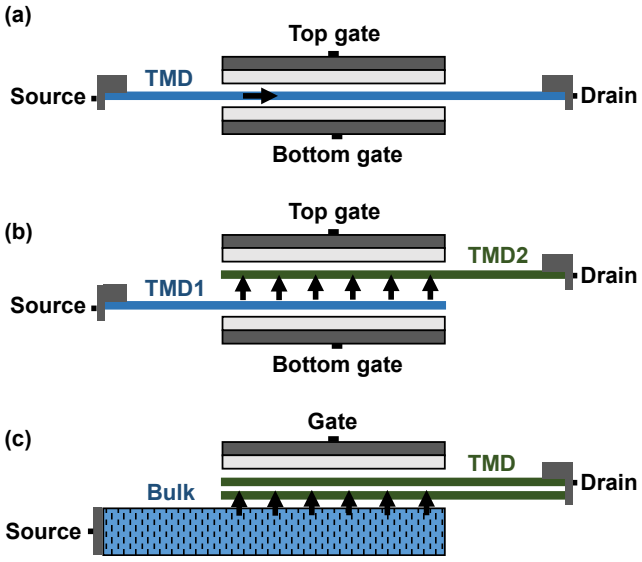

Figure 2.16: TMD TFET configurations. (a) homojunction TMD TFET [85, 86, 87], (b) 2D heterojunction TMD TFET, also called a Thin-TFET [88] and (c) 3D-2D heterojunction bilayer TMD TFET, also called ATLAS-TFET [72]. The arrows indicate the direction of the tunneling paths.

bandgap can be introduced by symmetry-breaking operations such as patterning the graphene into nanoribbons, or by stacking two layers and applying an electric field [90]. This allows the bandgap to be tuned, either by controlling the size of the ribbon or the strength of the electric field.

Transition metal dichalcogenides (TMD) form an alternative to graphene in 2D TFET applications. TMDs are a class of 2D materials with chemical formula MX_2, with M a transition metal and X a chalcogenide. In contrast to graphene, TMDs do exhibit a variety of relatively large bandgaps. They can be used as the channel material of an atomically thin TFET [85, 86, 87] (see Fig. 2.16(a)), or layers of different TMDs can be stacked as a heterostructure, bonded by Van der Waals interactions. Tunneling then occurs in the overlap region of the layers, perpendicular to the gate (see Fig. 2.16(b)). Such 2D heterostructure TFETs are predicted to combine high I_{ON} and low I_{OFF} in the same way as bulk heterostructures [88]. TMDs can also be combined with a bulk semiconductor, such as MoS_2 on Ge [72] (see Fig. 2.16(c)). Challenges that remain for 2D materials are the development of large area defect-free production, good ohmic contacts and techniques to obtain high doping levels [89, 91]. As a result of these challenges, only few experimental 2D TFETs with sub-60 mV/dec SS have been realized (see Fig. 2.28).

2.8 Dopant pockets

To improve SS and/or I_{ON}, another design option is to use dopant pockets. A dopant pocket is a localized region of doping, which locally modifies the device electrostatics. In a TFET, they can be used both in point tunneling (Section 2.8.1) and line tunneling (Section 2.8.2) configurations.

2.8.1 Pocketed pointTFET

In a pocketed pointTFET, a counterdoped pocket is added at the interface between source and channel to improve both I_{ON} and SS (see Fig. 2.17(a)). This forms a p-n-i-n doping profile in an nTFET, and a n-p-i-p profile for a pTFET. Compared to a p-i junction, the built-in field at a p-n junction is larger, corresponding to a stronger built-in band bending and hence an onset at lower V_{GS} and the potential for a higher I_{ON} [13, 92, 93, 94]. As shown in the band diagrams of Fig. 2.18, for a properly designed pocket, the enhanced band bending induces an abrupt transition from long to very short tunnel paths at tunneling onset, compared to more gradual shortening seen in a standard p-i-n/n-i-p

configuration. In the transfer characteristics this results in an improvement of SS, as shown in Fig. 2.19.

An additional effect of the pocket is that it reduces the impact of FIQC on the pointTFET performance. The strong built-in electric field at the tunnel junction forces the tunnel paths in a direction more parallel to the gate and induces tunneling before a strong band bending towards the gate, causing FIQC, is induced. The result is a shift of the BTBT onset to lower V_{GS} and a reduction

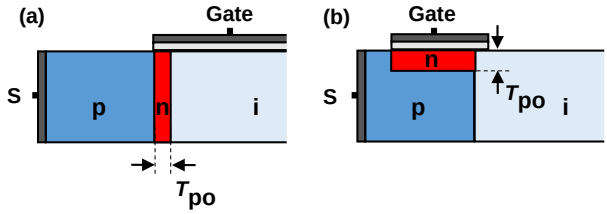

Figure 2.17: Source-channel region of a pocketed (a) point tunneling and (b) line tunneling TFET configuration.

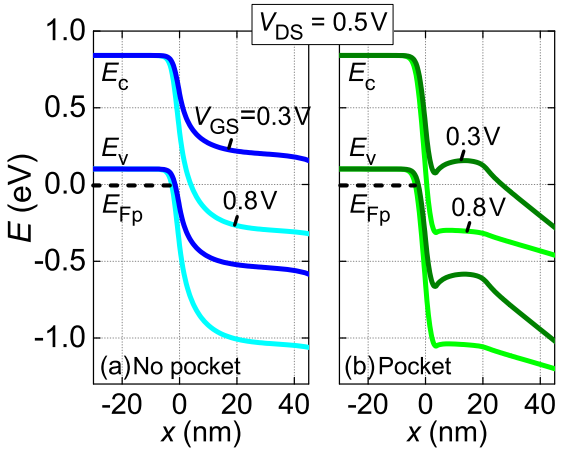

Figure 2.18: Energy band diagrams of a $In_{0.53}Ga_{0.47}As$ TFET in the OFF (dark lines) and ON-state (light lines). V_{GS} is respectively 0.3 V and 0.8 V. (a) A no-pocket configuration and (b) a configuration with a 4 nm thick pocket at the source-channel junction as in Fig. 2.17(a). The source and pocket doping is $5x10^{19}cm^{-3}$. The EOT is 0.6 nm, with a T_{body} of 20 nm. The dashed line represents the quasi Fermi level for the holes in the source (E_{Fp}).

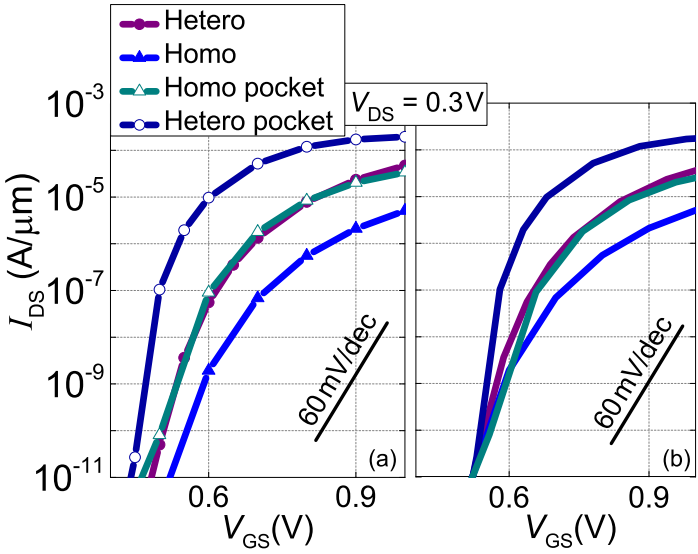

Figure 2.19: QM simulated transfer characteristics of Fig. 2.14 compared to the transfer characteristics of the same configurations with a counterdoped pocket of 3 nm. The TFET is a double gate configuration with an EOT of 0.6 nm, a T_{body} of 10 nm and a source and pocket doping of $5\mathrm{x}10^{19}\mathrm{cm}^{-3}$. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1\mathrm{x}10^{-11}\mathrm{A}/\mu\mathrm{m}$ coincides. This shift can be accomplished by a proper choice of gate work function.

of the stretching effect on the SS, discussed in Section 2.4.1.

An optimum exists for the pocket thickness (T_{po}, see Fig. 2.17(a)). The strength of the built-in field, and consequently the improvement in I_{ON}, increases with T_{po} (see Fig. 2.20(a)). However, if the pocket is so thick that the concentration of free carriers in the pocket becomes large (close to the doping level in the source) and is larger than the concentration in the channel at the onset of tunneling, the abrupt onset is deteriorated by a potential barrier in the channel [13] (see Fig. 2.20(b)). At the first moment of cross-over between valence and conduction band, a dip is present in the energy bands at the tunnel junction, followed by a bump in the channel. For such a configuration, a further increase in V_{GS} only marginally goes toward increasing the band bending at the tunnel junction, but is consumed largely by a further increase of the carrier concentration in the pocket. The structure is then effectively a p-n diode in series with an n-i-n MOSFET. A good estimate for the optimal T_{po} is the depletion width at the tunnel junction [13, 92]. Note that for configurations with a large T_{body}, a thick

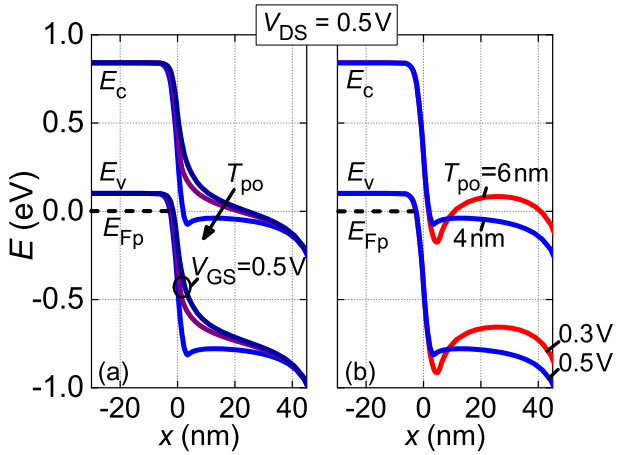

Figure 2.20: Energy band diagrams of an $In_{0.53}Ga_{0.47}As$ TFET around onset of BTBT. The source and pocket doping is $5x10^{19}cm^{-3}$. The EOT is 0.6 nm, with a T_{body} of 20 nm. (a) Impact of increasing T_{po} (0 nm, 2 nm and 4 nm) for a constant V_{GS} of 0.5 V, showing an increasing electric field at the tunnel junction. (b) Comparison of an optimal T_{po} of 4 nm to a larger T_{po} of 6 nm at onset, with the latter showing a potential bump in the channel. V_{GS} is respectively 0.5 V and 0.3 V. The dashed line represents the quasi Fermi level for the holes in the source (E_{Fp}).

pocket can cause an uncontrolled tunneling current through the body of the device.

2.8.2 Pocketed lineTFET

In a pocketed lineTFET, a pocket is introduced to mitigate the impact of FIQC and to reduce the sensitivity to the gate-channel overlap (see Section 2.4.2). The pocket is located in the source region, underneath the gate-source overlap and adjacent to the source-channel junction (see Fig. 2.17(b)). Similar to the pocketed point TFET, the doping type of the pocket is opposite to that of the source.

The effect on the FIQC underneath the gate dielectric is shown by the energy bands of Fig. 2.21: the triangular well is rounded off and becomes wider for increasing T_{po}. The rounding reduces the FIQC, such that the subband quantization levels decrease in energy. The onset of vertical BTBT, determined by the first quantized level, therefore shifts to a lower V_{GS}. The resulting onset

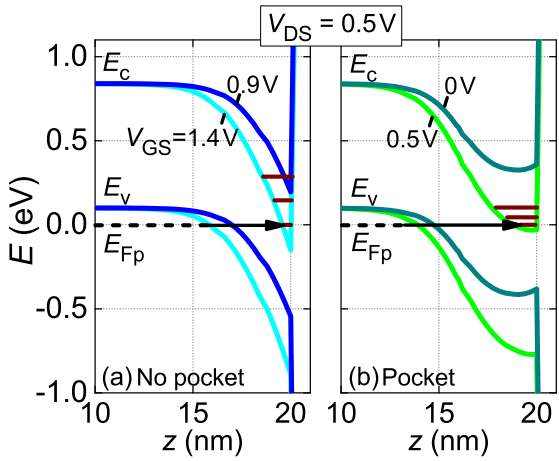

Figure 2.21: Energy band diagrams of an $In_{0.53}Ga_{0.47}As$ lineTFET in the OFF (dark lines) and ON-state (light lines) along a vertical cutline through the center of the gate. The source and pocket doping is $5x10^{19}cm^{-3}$. The EOT is 0.6 nm, with a T_{body} of 20 nm. (a) A no-pocket configuration for a V_{GS} of 0.9 V and 1.4 V and (b) a configuration with a 4 nm thick pocket underneath the gate as in Fig. 2.17(b) for a V_{GS} of 0 V and 0.5 V. Indicated are estimations of the first three quantized energy levels inside the potential well for the highest V_{GS}. The dashed line represents the quasi Fermi level for the holes in the source (E_{Fp}).

voltage V_{onset} can be calculated analytically as [95]:

$$V_{onset} = \frac{GWF - E_{EA} + E_{sub}}{q} - \frac{q(N_{po} + N_{source})T_{po}}{C_{ox}} + \frac{1}{C_{ox}} \times \sqrt{2q\epsilon_s N_{source}\left[\frac{E_G + E_{sub}}{q} + \frac{q(N_{po} + 2N_{source})T_{po}^2}{2\epsilon_s}\right] - q^2 N_{source}^2 T_{po}^2} \tag{2.27}$$

provided the pocket is fully depleted, with GWF the work function of the gate, E_{EA} the semiconductor electron affinity, N_{po} and N_{source} respectively the doping level of the pocket and the source and E_{sub} the energy level of the first subband in the potential well underneath the gate, which can be solved from a 1D Schrödinger equation. Eq. (2.27) shows that by adjusting the thickness and doping of the pocket for a given GWF, the FIQC-induced shift in V_{onset} seen in the no-pocket case can be removed. Note that for materials which are more prone to FIQC, i.e. those with a small effective mass and hence a large E_{sub}, the required T_{po} or N_{po} to compensate for the FIQC shift are larger.

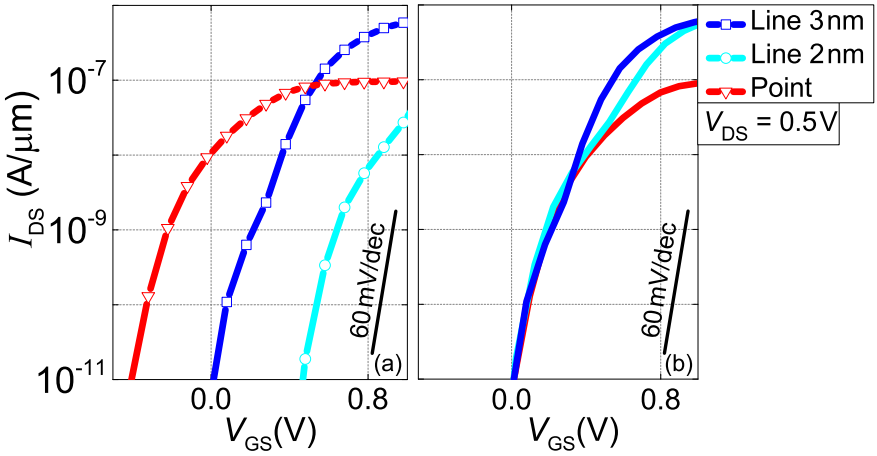

Figure 2.22: QM simulated transfer characteristics of a Si pocketed pointTFET with a T_{po} of 4 nm and pocketed lineTFET with two different T_{po}. The TFETs are double gate configurations with an EOT of 0.6 nm, a T_{body} of 20 nm and a source doping of $1\mathrm{x}10^{20}\mathrm{cm}^{-3}$. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1\mathrm{x}10^{-11}\mathrm{A}/\mu\mathrm{m}$ coincides. This shift can be accomplished by a proper choice of gate work function.

In materials like Si, which have two distinct subband ladders in the well underneath the gate, the decreased FIQC has the added benefit of reducing the relative shift between the subband levels. The corresponding kink in the SS which was discussed in Section 2.4.2 is thereby also reduced, as shown in Fig. 2.22, where two pocketed lineTFETs are compared to a pocketed pointTFET configuration. For a thicker pocket, V_{onset} is reduced, while the relative shift of the two current components is decreased. Note that this kink is not present in a III-V configuration with an isotropic conduction band effective mass.

The presence of the counterdoped pocket also reduces the unwanted lateral point tunneling component by blocking the shortest lateral tunneling paths closest to the gate, since the source-channel p-i junction is locally replaced by a n-i junction. Combined with the earlier onset of vertical BTBT, the overall impact of the lateral component on the transfer characteristics is thus diminished, removing also the variability to the gate alignment.

The comparison of transfer characteristics in Fig. 2.22 shows that the pocketed lineTFET outperforms the pocketed pointTFET in a Si configuration in terms of I_{ON} and SS [13].

2.9 Gate stack configuration

As discussed in Section 2.6, the gate control is an essential factor in the BTBT efficiency. Gate control is characterized by two aspects. The first aspect is the portion of the gate voltage observable in the semiconductor, after the voltage drop over the oxide. This is directly related to the EOT. A second aspect is the degree in which a given variation in gate voltage moves the equipotential lines in the device closer together and hence increases the electric field locally. By tuning both aspects, the gate control at a tunnel junction can be enhanced or decreased.

To improve TFET performance for conventional logic applications, it is advantageous to enhance the gate control at the source-channel tunnel junction (Section 2.9.1), and to decrease it at the channel-drain junction (Section 2.9.2). The former improves the TFET SS and I_{ON}, while the latter reduces the ambipolar current and C_{GD}. In this section, gate configurations addressing both requirements are discussed.

2.9.1 Improving source-channel BTBT

A first way to improve the gate control at the source-channel junction is to decrease the EOT, which increases the electrostatic potential drop in the semiconductor. Like a MOSFET, the TFET performance therefore benefits from the introduction of high-k materials as the gate dielectric [3, 96, 97].

A second way to improve the gate control at the source-channel junction is to go from a single gate (SG) to a multi-gate (MuG) to a gate-all-around (GAA) configuration [98, 99] (see Fig. 2.23), an approach which is also applied to MOSFET. This requires a transition from a planar to a protruding structure. A multi-gate configuration can be implemented as a finFET, where the gate covers three sides of the device body. Going further, the TFET body can be realized as a horizontal or vertical nanowire (NW), with the gate wrapped around.

The tighter gate control can be seen when comparing the electrostatic profile of a single gate to that of a double gate configuration [98] (see Fig. 2.24). When a second gate is added, the equipotential lines are curved more strongly. A consequence is that the electric field close to the source region is enhanced and hence the tunnel path lengths are reduced, improving I_{ON}. The electric field is now increasing towards the source region. Therefore, a given increase in V_{GS} results in a larger increase in electric field than in the single gate configuration, because the electric field now increases both due to a larger total voltage drop between source and gate, and because the tunneling moves closer to the source

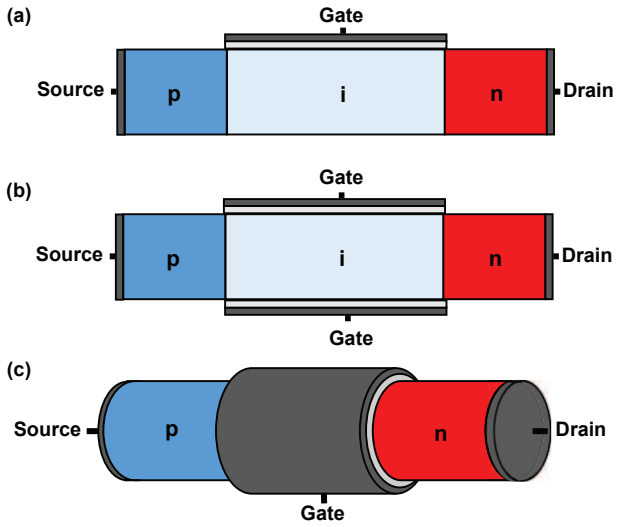

Figure 2.23: (a) Single gate, (b) double gate and (c) gate-all-around TFET configuration.

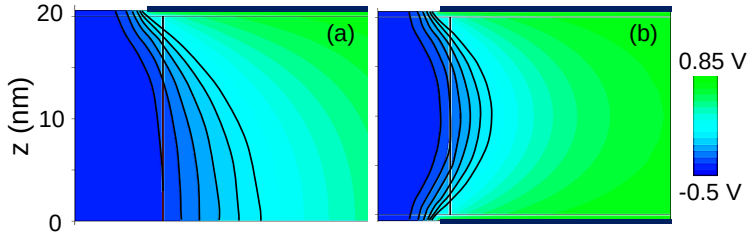

Figure 2.24: Electrostatic potential profiles at the source-channel junction for a (a) single gate and (b) double gate $In_{0.53}Ga_{0.47}As$ p-i-n TFET at a V_{GS} of 1 V and a V_{DS} of 0.5 V. Six contour lines have been highlighted in each configuration.

where the equipotential lines are more closely spaced, and the electric field is higher. The result is a beneficial impact on the SS. This effect is enhanced further for a MuG or GAA structure as illustrated in Fig. 2.25(a).

Further scaling of the body thickness has an additional beneficial impact on the gate control. As the gates move closer together, the curvature of the

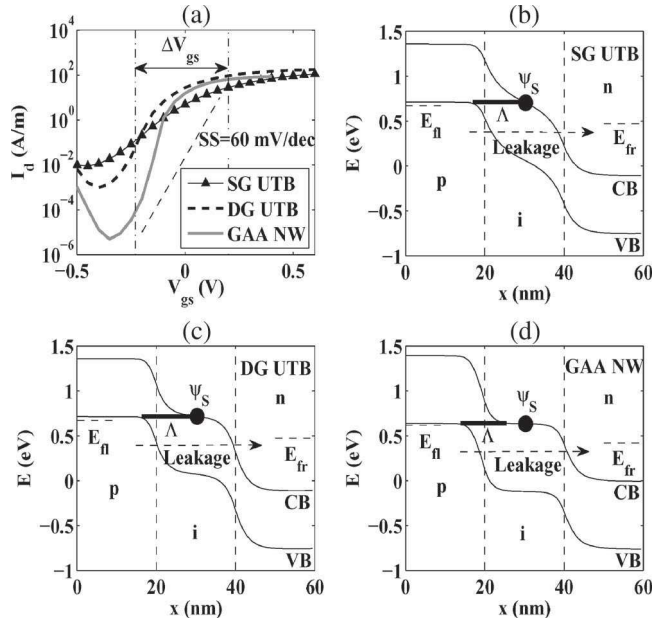

Figure 2.25: (a) QM simulated transfer characteristics at $V_{\rm DS} = 0.2$ V of InAs (line with triangles) SG ultra thin body (UTB), (dashed line) DG UTB, and (gray line) GAA NW devices. $T_{\rm body}$ or diameter is 6 nm. The NW current is normalized with its diameter. The OFF-to-ON gate voltage swing Δ-$V_{\rm GS}$ is indicated. (b) CB and VB edges along the transport direction x of the 6 nm p-i-n SG UTB device at the onset gate voltage. Λ indicates the tunneling barrier width, $\psi_{\rm S}$ is the electrostatic potential in the middle of the gate, and $E_{\rm fl}$ and $E_{\rm fr}$ denote the source and drain Fermi levels, respectively. (c) Same as (b) for the 6 nm DG UTB device. (d) Same as (b) and (c) for the 6 nm GAA NW device. Current leakage paths are indicated by dashed arrows in (b), (c), and (d)

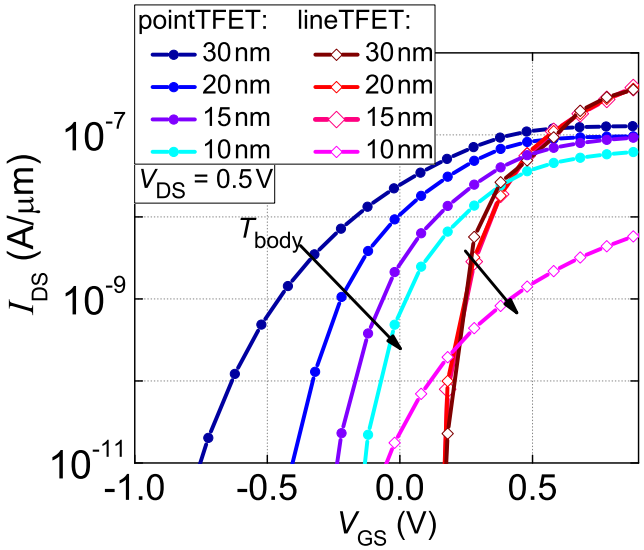

Figure 2.26: QM simulated transfer characteristics of Si pocketed point and lineTFETs for varying T_{body}. The configurations are as shown in Fig. 2.17, but have a double gate. The gate-source overlap for the 10 nm body pointTFET configuration is removed. For the lineTFET, the length of the gate-source overlap and the pocket are 20 nm. The EOT is 0.6 nm. Source and pocket doping are $1\text{x}10^{20}\text{cm}^{-3}$.

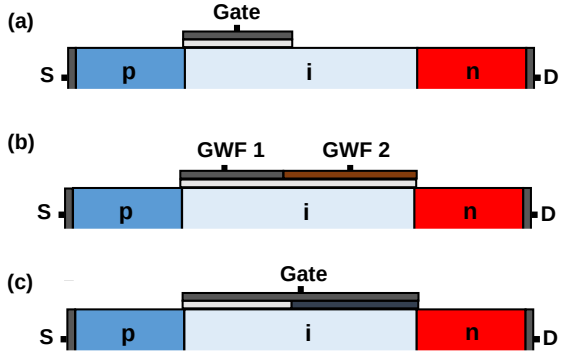

Figure 2.27: Gate stack configurations for a p-i-n TFET aimed at reducing ambipolar behavior. (a) Short gate, (b) dual metal work functions and (c) dual EOT dielectric.

equipotential lines becomes more pronounced, and the electric field hence increases. As explained in the previous paragraph, this benefits the SS of a pointTFET as illustrated in Fig. 2.26 for a pocketed point TFET. For a double gate lineTFET, the body scaling does not benefit performance above a certain minimum T_{body}, as the gate depletes the source locally underneath the gate and tunneling starts from close to the undepleted source region (see Fig. 2.21), which is at source voltage. Hence, the two depleted source regions underneath the two gates are disconnected in the lineTFET.

Specific care must be taken when scaling the body thickness of lineTFET or pointTFET configurations, to prevent excessive depletion of the source region [13, 100]. Underneath the gate-source overlap, a depletion region extends into the source, with a maximal extent in the ON-state (as explained in Section 2.3.1). If the body of the device is too thin, the depletion regions of different gates touch even before the device reaches the ON-state, resulting in depletion of the source over the full body thickness. For a lineTFET, this means no more tunnel paths perpendicular to the gate are available. Only much longer non-orthogonal tunnel paths are available and the current drops accordingly (see Fig. 2.26). Since the lineTFET operation depends on the gate-source overlap, the body thickness can therefore not be scaled below a certain critical thickness. This thickness is determined by the sum of the depletion lengths underneath the gate-source overlaps in the ON-state. The depletion lengths themselves are determined by the doping level of the source and of the pockets, if present. In a pointTFET, depletion over the full body thickness increases the tunnel path lengths, resulting in a drop in current for high V_{GS}. In this case, full depletion of the source can be avoided by reducing or removing the gate-source overlap for thin body thicknesses.

2.9.2 Reducing channel-drain BTBT and C_{GD}

The ambipolar current can be reduced by introducing a gate-drain underlap [78] (see Fig. 2.27(a)). The underlap decreases the gate control on the channel-drain junction. Since the desired BTBT takes place locally at the source-channel junction, the gate does not have to cover the full channel as in a MOSFET to enforce a particular band bending and hence tunnel path length at the source-channel junction. The underlap increases the distance between the gate and the channel drain-junction and consequently reduces the BTBT efficiency at the drain. The short gate has the added benefit of reducing C_{GD}, which is responsible for the Miller effect (see Section 2.11.3). Shortening of the gate can proceed down to the length at which it loses control over the source-channel junction (typically lengths of 10 nm and below). Care has to be taken as well that the resistance in the ungated channel region does not surpass the resistance

of the tunnel barrier, to avoid an observable reduction in the desired TFET current.

Instead of removing part of the gate, an alternative is to make the gate heterogeneous, for instance by introducing two different metals, each with a different work function [101] (see Fig. 2.27(b)). This is also called a dual material gate TFET. Although both metals are biased at the same V_{GS}, the work function at each of the tunnel junctions (source-channel and channel-drain) determines the amount of local band bending and hence the local tunnel barrier. The two work functions can be optimized separately: at the source-channel junction for a minimal tunnel barrier in the ON-state and at the channel-drain junction for a maximal tunnel barrier in the OFF-state.

The same can be achieved by reducing the EOT only at the source-channel junction, while retaining a large EOT at the channel-drain interface [102] (see Fig. 2.27(c)). This can be implemented with different physical oxide thicknesses, or with a high-k dielectric at the source side and a low-k dielectric at the drain.

2.10 Strain

Similar to a MOSFET, mechanical strain impacts TFET performance. This is a result of the effect strain has on the band structure (Section 2.10.1), an effect which can be incorporated in TFET models (Section 2.10.2). Strain can be externally applied as a performance booster, and is also intrinsically present at a lattice-mismatched heterojunction (Section 2.10.3).

2.10.1 Impact of strain on the band structure

Mechanical strain can have an important effect on the band structure of a semiconductor, since it alters the bond lengths and changes the crystal symmetry [103]. Firstly, strain warps the curvature of the energy bands, and hence influences the effective masses. This effect is exploited in MOSFETs to increase carrier mobility, by choosing a strain configuration such that the effective masses in the transport direction are decreased [104]. Secondly, strain shifts the relative positions of valley edges, such that the bandgap changes. Thirdly, as a result of broken symmetries, strain lifts certain energy band degeneracies, which modifies the DOS. Most notably, strain can lift the degeneracy between the heavy and light hole valence band.

2.10.2 Strain modeling

A common approximate model to study the effect of strain on the band structure is the deformation potential model [105]. The strain due to the deformation is assumed to be small and gradually varying, such that the effect on the energy band edges can be treated as if it were a small applied electrostatic potential. The shift in the band edges then varies linearly with the strain components:

$$\Delta E_n = \sum_{ij} \Xi_{ij} \varepsilon_{ij} \tag{2.28}$$

with E_n the band edge of band n, ε_{ij} the elements of the strain tensor and Ξ_{ij} the so-called deformation potential constants, available in literature. The deformation potential model can be directly applied in semiclassical effective mass-based simulators as a correction on the band edges.

Strain can be introduced into more complex band structure models as well. In the **k**·**p**-model, strain can be incorporated by the addition of a strain interaction term as proposed by Pikus and Bir [106], which includes the deformation potential constants of Eq. (2.28). In the TB approach, the integrals of Eq. (2.21) are adjusted based on the strain-induced change in bond lengths and angles [107]. The introduction of strain in these models accounts for both the shift in band edge energies and the change in effective masses. Strained band structure models can be used in fully quantum mechanical simulations, but are also useful to extract band edge energies and effective masses, which serve as inputs to semiclassical models [108].

2.10.3 Strain in TFET

Since TFET operation depends heavily on the band structure, strain affects the TFET performance. Strain can be induced intentionally like in a MOSFET, e.g. by external stressors, to improve I_{ON} by reducing E_{G} and m_{R} [17] (see Section 2.6.1). Similarly, the shifting of the band edges due to strain can induce a more favorable band alignment at a heterojunction [109] (see Section 2.7.3). Specifically in indirect materials, the relative shift of conduction band valleys can render the material direct [110]. This is of particular interest in Ge, where the difference between the indirect and direct bandgaps is small.

Strain can also be inherently present in a TFET, e.g. at the heterojunction of two lattice mismatched materials. The strain profile in this case is very non-uniform, making it difficult to predict the exact impact on the TFET performance.

2.11 Circuit considerations

Certain TFET features require additional care when designing circuits. The most important are the ambipolar effect (Section 2.11.1), the asymmetrical structure (Section 2.11.2), the Miller capacitance (Section 2.11.3) and the superlinear onset in the output characteristics (Section 2.11.4).

2.11.1 Ambipolar effect

The TFET ambipolarity can result in an increased switching leakage in a TFET inverter, but can just as well be used to design more compact digital and analog circuits if the positive and negative switching of the gate voltage to turn on the tunneling current is exploited [111, 112]. For good ambipolar behavior in these applications, both the source-channel and channel-drain junction have to be optimized to obtain similar I_{ON} and SS. This stands in contrast to conventional logic applications, in which the channel-drain tunneling is suppressed (Section 2.9.2). In digital circuits, the ambipolarity can be used to create more complex logic gates with fewer components [111]. These ambipolar gates have control inputs that determine the bias point, and hence the polarity of the devices. In analog applications, an ambipolar amplifier for example, exhibits both positive and negative small-signal gain, depending on the bias point [112]. This allows it to function as both a common-source and common-drain amplifier, without changing the physical implementation.

2.11.2 Asymmetrical structure

Unlike a MOSFET, the TFET has an asymmetrical source and drain doping and therefore behaves significantly different for a positive or negative V_{DS} polarity [113]. As the p-i-n diode is reverse biased in normal TFET operation, the sign of V_{DS} is fixed to positive for nTFET and negative for pTFET. When properly designed, the dominant current is BTBT-based with leakage mechanisms as described in Section 2.6.3. Reversing the sign of V_{DS} brings the p-i-n diode in forward bias. At small forward bias, the current is typically negligible compared to the current at similar reverse bias, except for designs with sufficiently high source doping such that (Esaki-) tunneling becomes significant [114]. Even then, a drop in current is typically present upon increasing V_{DS} as the tunneling decreases, before the forward bias diode diffusion current reaches sufficiently high values to become dominant.

The asymmetrical TFET conduction can be a concern for Static Random Access Memory (SRAM) circuits. In standard 6-transistor (6T) SRAM cells, the two access transistors are required to conduct current in two directions. The asymmetrical conduction leads to unacceptable read and write static noise margins and hence instability of the SRAM cell [115]. Alternative SRAM cell designs with increased number of transistors have been proposed to circumvent this issue [115, 116].

Weak conduction in one direction also prevents the discharging of transient noise signals that appear on a circuit node through capacitive coupling [117]. These signals can at times become very large and significantly surpass V_{DD} and therefore be detrimental to device reliability. Also the circuit timing suffers from these excessive noise signals. To alleviate this issue, the forward Esaki tunneling current present at low V_{DS} can be used to evacuate the excess charge. This requires a sufficiently high degeneracy in the source.

Several other logic circuits, like the basic inverter, are mostly unaffected by the asymmetry. It is even possible for specific basic circuits to exploit the unidirectionality to redesign the circuit with fewer transistors than in conventional CMOS, e.g. the pass-gate multiplexor [117].

2.11.3 Miller capacitance

A high C_{GD} creates an enhanced Miller effect, which results in over-and undershoots in the transient behavior of an inverter. The peak voltages V_{P} of the over- and undershoots can be expressed as [118]:

$$V_{\mathrm{P}} = \frac{C_{\mathrm{M}}}{C_{\mathrm{M}} + C_{\mathrm{L}}} V_{\mathrm{DD}} \tag{2.29}$$

with C_{L} the load capacitance of the inverter and with C_{M} comprising the C_{GD} of both nTFET and pTFET of the inverter. It is possible to decrease V_{P} by increasing C_{L}, but this has the downside of also increasing the inverter fall time delay. Care must therefore be taken in the device design to keep C_{GD} low in the first place.

The gate capacitance division in gate-source capacitance C_{GS} and gate-drain capacitance C_{GD} is different from the MOSFET-case, because charge injection from the source requires a tunneling event, while charge injection from the drain is via thermal injection, similar to the MOSFET-case. Therefore, when the tunneling efficiency at the source side is comparatively low, and in the absence of a strong SRH process, the carriers in the channel are supplied by the drain and the full channel determines C_{GD}. C_{GS} is then determined by the depletion of the highly-doped source region, hence resulting in a decreasing capacitance

with increasing V_{GS} (nTFET). The channel charge is determined by the position of the drain Fermi level and the conduction band in the channel (nTFET). A channel material with a lower DOS in the conduction band (nTFET) hence results in a smaller charge build-up in the channel and the corresponding C_{GD} is also smaller [118]. Despite the fact that $C_g \approx C_{GD}$ if the tunneling efficiency at the source is low, the C_{GD} can hence be limited if the TFET is properly designed. Also a gate-drain underlap can help to reduce the Miller effect [119].

In cases where the tunneling efficiency is high and the device is fully in the ON-state, a significant amount of charge in the channel is supplied by the source. Like a MOSFET, the TFET then exhibits a linear region, in which charge is contributed equally by source and drain, and a saturation (pinch-off) regime, in which charge is uniquely supplied by the source. The gate capacitance is divided accordingly between the contacts. Improving the source tunneling efficiency is therefore also beneficial in reducing the Miller effect.

2.11.4 Superlinear onset

As discussed in Section 2.3.3, the TFET output characteristics can have a superlinear onset [38, 39, 40]. The small current at low V_{DS} delays the last part of the charging or discharging of the load capacitance in a TFET inverter. This increases the inverter rise and fall times and hence increases the total circuit delay. Additionally, recovery from transient noise signals is less efficient when the output characteristics are superlinear [120]. As suggested in Section 2.3.3, the superlinear onset can be reduced by improving the BTBT efficiency at small V_{DS}.

2.12 Experimental status

After describing the device physics and architecture options in previous sections, this section is intended to give an overview of experimental realizations. The current status of experimental TFET transfer characteristics is summarized in Fig. 2.28 [72, 9, 36, 69, 71, 73, 94, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135]. The curves have been selected to have a point-SS lower than 70 mV/dec, which means they are either close to or below the ideal MOSFET limit of 60 mV/dec for a given voltage range. A variety of experimental techniques is used to fabricate these structures [27]. Some configurations can be seen to obtain a sub-60 mV/dec SS, demonstrating the steep switching potential of TFET. None of the displayed curves, however, attains all requirements for SS, I_{ON}, I_{OFF} and I_{60} as specified in Section 2.6,

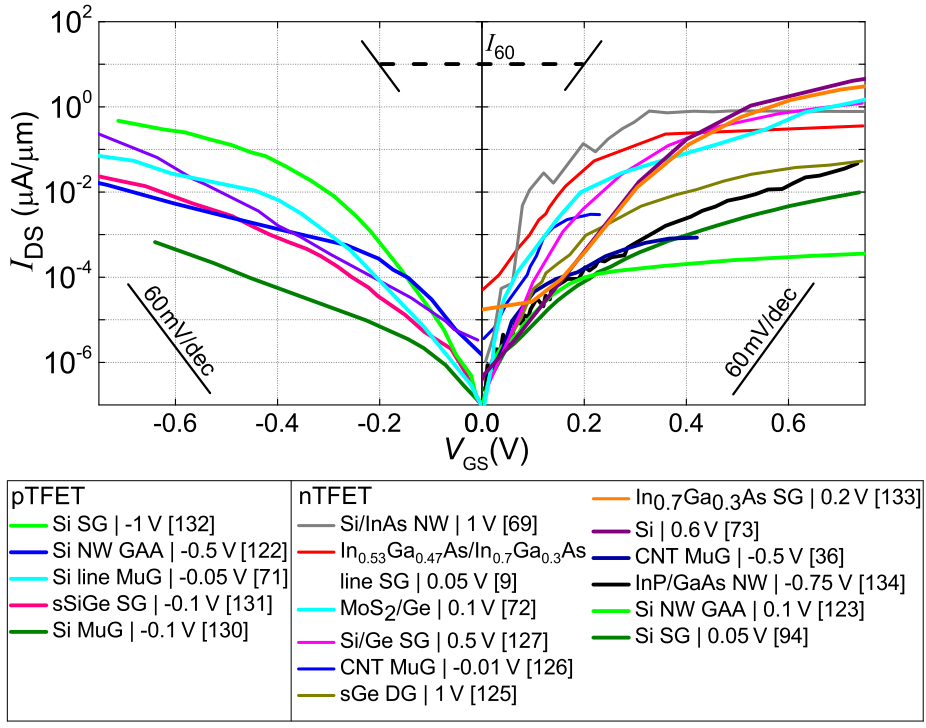

Figure 2.28: Experimental transfer characteristics in literature of nTFET and pTFET. Only curves with an SS lower than 70 mV/dec have been included. The legend includes the material system (s-prefix stands for strained), gate configuration and V_{DS}. The legend entries are ordered according to the intersection of the corresponding curves at a V_{GS} of -0.2 V and 0.2 V. Also indicated is the target value for I_{60} mentioned in Section 2.6.2. This figure is an updated version of Fig. 2 of Lu *et al.* [26].

and which are required to be a viable replacement for MOSFET. The comparison between nTFET and pTFET shows that fewer experimental realizations of pTFET exist, with a performance generally inferior to nTFET. The reasons behind this discrepancy are discussed in Section 2.6.2. It is clear from Fig. 2.28 that a gap still needs to be closed between the promising theoretical predictions and actual experiment. Note, however, that some configurations which are promising in simulations, confined III-V heterostructure NW, are challenging to fabricate, and have not yet been experimentally realized.

2.13 Conclusions and outlook

In this article, the TFET was discussed from the perspective of device physics. The basic operating principle based on BTBT enables the TFET to obtain a sub-60 mV/dec SS at room temperature. This makes it a promising candidate to circumvent the increasing power issues in scaled MOSFET technologies. The main challenges for TFET are to obtain a low average SS, a high I_{ON} and low I_{OFF}, or stated differently: a high I_{60}, and this for both nTFET and pTFET. Accurate predictive models are required to meet these challenges, with quantum mechanical simulators complementing more and more the semiclassical models typically used for MOSFET. These models can help in choosing the optimal material system, with alternative group IV materials, III-V materials and 2D materials proving interesting alternatives for Si. Combining materials in a heterostructure provides an additional degree of freedom in the device design. Other performance improvements can be gained from dopant pockets or from an optimization of the gate stack configuration. The presence of strain can have a beneficial effect on the device performance. Finally, several specific characteristics of the TFET operation at a device level also have consequences for its use in circuits, suggesting the need for co-optimization of device and circuit. The gap between theoretical predictions and experimental realizations today shows that many challenges are still to be overcome, before the TFET's appearance in low-power products.

Chapter 3

Formalism

In the previous chapter, we introduced the TFET and showed that to meet its challenges, a move towards III-V heterostructure implementations is expected. We also saw that this move requires adequate and computationally efficient modeling approaches to explore the design space of new materials and configurations. We discussed various SC and QM modeling frameworks. In this chapter, we select and tailor a suitable simulation formalism based on the requirements set by our application. We select the EF band structure model and combine it with the QTBM in a ballistic WF approach. The resulting system is solved numerically with a FD scheme. This chapter presents a general derivation, valid for an arbitrary number of bands, and forms the basis for the subsequent chapters, which discuss particularities of the implementation of specific band models required to simulate TFET configurations of increasing complexity.

The structure of this chapter is as follows. First, we motivate our formalism choice based on a set of requirements (Section 3.1). We then derive a system of 2D heterostructure EF equations, starting from the one-electron time independent Schrödinger equation (Section 3.2). We construct boundary conditions for this EF system using the quantum transmitting boundary method (QTBM) (Section 3.3). Next, we propose a discretization that allows us to solve the EF system numerically (Section 3.4). Finally, we derive expressions to calculate the transmission probability and current density in the device from the obtained EFs (Section 3.5). The chapter ends with a conclusion (Section 3.6).

3.1 Requirements and model choice

3.1.1 Requirements

The goal of this thesis is to efficiently make predictions for the performance of heterostructure TFETs. In Section 2.5, we discussed the options that we have when developing a TFET modeling approach: we can take a semiclassical (SC) or a quantum mechanical (QM) physical model, a $\mathbf{k}\cdot\mathbf{p}$/envelope function (EF) or tight-binding (TB) band structure model, an non-equilibrium Green's function (NEGF) or wave function (WF) solution method and a numerical discretization with finite differences (FD), finite elements (FE) or a spectral method. To choose between all of these options, we first define a set of requirements for our modeling approach:

1. **BTBT modeling:**
   BTBT determines the current flow in a TFET. The description of BTBT requires a model that includes multiple bands between which charge carriers (electrons and holes) can transition.

2. **2D potential profiles:**
   Since the TFET is a device with three terminals, the electrostatic potential has a distinctly two-dimensional character. The device is also finite in the lateral direction (which we will refer to as "confined"). The model should reflect the effects of the 2D potential profile and confinement, like FIQC and SIQC, on the carrier distribution and transport.

3. **Material change:**
   As discussed in Chapter 2, heterostructures are promising to improve TFET performance. The chosen band structure model should therefore remain valid in each of the separate materials of a heterostructure. Additionally, reflections at the heterojunction interface should be taken into account [52].

4. **Computational efficiency:**
   Computation power is not infinite, but bounded by practical resources. The chosen model and numerical solution method should allow for the simulation of realistic bulk TFET devices within the available computational capabilities.

3.1.2 Model choices

We now motivate our choices for a particular physical model, band structure model, solution method, boundary conditions and numerical implementation based on the requirements in the previous section.

For the physical model, requirements 2 and 3 rule out SC models, and require a QM approach. SC models do not capture confinement effects, such as FIQC or SIQC, which has been shown in group IV material systems to lead to overly optimistic TFET performance predictions [1] (see also Section 2.5.3). SC models also do not include reflections at a heterojunction and therefore overestimate the transmission probability [18].

For the choice of band structure model, requirement 4 rules out ab-initio or atomistic methods such as density functional theory (DFT) [136], the TB method [55] and methods based on pseudopotentials [137], as they require large computational resources. We therefore choose the **k**·**p**-based EF formalism as developed by Burt [54]. As discussed in Section 2.5.2, the EF formalism is a fully QM approach, in which a system of differential equations is constructed based on an EF expansion of the wave function. It is a continuum method, based on an extension of the **k**·**p**-method for bulk semiconductor band structures to cases in which an external potential is applied. The continuum character limits the computational burden compared to atomistic or ab-initio methods, which satisfies requirement 4. The EF formalism can be implemented with various numbers of bands depending on the application, ranging from a one-band implementation, which is equivalent to the effective mass method, to thirty and even fourty-band methods that capture the full first Brillouin zone. Tunneling between these bands is captured by coupling terms, which can be calibrated to experimental data on effective masses and energetic gaps. Requirement 1 is therefore satisfied by including at least two bands. For requirement 3, the validity of the Burt EF formalism has been rigorously extended by Van de Put *et al.* to heterostructures [52] and it has been shown that reflections at the heterojunction are captured [18].

For the solution method, we combine the EF formalism with the QTBM to impose open boundary conditions at the source and drain contacts (assuming a constant potential there, see Section 2.5.2), and solve the EF equations in a WF approach. A prominent alternative approach is the NEGF method discussed in Section 2.5.2, but since we do not consider scattering, NEGF does not have an added value over a WF approach. On the contrary, NEGF would result in a higher computational burden, since it requires the inversion of a large matrix (see Eqs. (2.22) and (2.23)) instead of the solution of a linear system as in the WF approach. A similar argument can be made against solution methods

Table 3.1: Model choices.

Physical model	Fully QM
Band structure model	$\mathbf{k}\cdot\mathbf{p}$/EF
Solution method	WF approach
Boundary conditions	QTBM
Numerical implementation	FD/Spectral method

based on Wigner distribution functions. Although these methods are very useful to study time dependent phenomena in the presence of various scattering mechanisms, they are also computationally very expensive [138]. This makes them unpractical for 2D multi-band simulations of extended devices.

Numerically, we discretize and solve the EF system with a FD scheme to satisfy requirement 2 without exceeding the computational requirement 4. Although FE methods are more flexible, the computational complexity makes them unfeasible for multi-band simulations. In Chapter 5, we keep the computational burden acceptable for band models with a large number of bands (more than two) by replacing the FD discretization in the confined direction with a spectral decomposition.

Based on these choices, summarized in Table 3.1, we now derive a system of EF equations, its boundary conditions, and expressions to calculate currents and transmission probabilities.

3.2 EF formalism

3.2.1 EF expansion

As we have briefly discussed in Section 2.5.2, the starting point to derive the EF system is the time-independent one-electron Schrödinger equation, which we repeat here:

$$\left[\frac{-\hbar^2}{2m_e}\nabla^2 + V_{\mathrm{e}}(\mathbf{r}) + V_{\mathrm{c}}(\mathbf{r})\right]\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{3.1}$$

where m_e is the free electron mass, $\hbar$ is the reduced Planck's constant, E is the total energy, $V_{\mathrm{e}}(\mathbf{r})$ is the potential energy that results from the external bias applied to the device contacts and $V_{\mathrm{c}}(\mathbf{r})$ is the internal crystal potential in the semiconductor crystal. In the case of a heterostructure, $V_{\mathrm{c}}(\mathbf{r})$ varies from one

material to the other, which can be modeled as a piecewise constant function:

$$V_{\mathrm{c}}(\mathbf{r}) = \sum_{l}^{L} \theta_l(\mathbf{r}) V_{\mathrm{c}}^{l}(\mathbf{r}) \tag{3.2}$$

where $V_{\mathrm{c}}^{l}(\mathbf{r})$ is the bulk crystal potential in material l and L is the number of material layers that make up the heterostructure. θ_l selects the layer l and is defined as a logical step function, which is only non-zero for $\mathbf{r}$ in the volume of layer l, denoted as Ω_l:

$$\theta_l(\mathbf{r}) = 0/1 \text{ if } \mathbf{r} \notin / \in \Omega_l \tag{3.3}$$

In the EF approach, the one-electron wave function ψ of Eq. (3.1) is decomposed into a set of basis functions U_n^k:

$$\psi(\mathbf{r}) = \sum_{n}^{N} F_n^k(\mathbf{r}) U_n^k(\mathbf{r}). \tag{3.4}$$

with F_n^k the envelope functions [54]. By construction, the EFs are slowly varying over a unit cell, such that they contain only Fourier components inside the first Brillouin zone. This means all strongly oscillating components of the wave function are contained in the basis functions. The set of basis functions can in principle be any complete and orthonormal basis set with the periodicity of the lattice. The number of basis functions N corresponds to the number of bands. N is infinite for an exact decomposition of the wave function, but is in practice determined by a trade-off between desired accuracy of the band structure and the computational burden.

In the EF approach that we take, the basis functions are eigenfunctions of the bulk Hamiltonian at the Γ-point, also called the zone-center. In the case of a heterostructure, they are eigenfunctions in only one of the materials, denoted as material k and indicated with a superscript in Eq. (3.4). In the other materials of the heterostructure, these basis functions are not eigenfunctions of the bulk Hamiltonian. However, a unitary transformation was developed by Van de Put *et al.* [52], which allows one set of basis functions to be used throughout the full heterostructure. We will make use of this transformation at a later stage of the derivation of the EF system.

3.2.2 Derivation heterostructure EF system

To derive the EF system, we now insert the EF expansion of Eq. (3.4) into Eq. (3.1) to obtain:

$$\begin{aligned}
&\frac{-\hbar^2}{2m_e}\sum_n \nabla^2 F_n^k(\mathbf{r})U_n^k(\mathbf{r}) - \frac{2\hbar^2}{2m_e}\sum_n \nabla F_n^k(\mathbf{r})\cdot\nabla U_n^k(\mathbf{r}) \\
&+\sum_n F_n^k(\mathbf{r})\left[\frac{-\hbar^2}{2m_e}\nabla^2 + V_{\mathrm{c}}(\mathbf{r})\right]U_n^k(\mathbf{r}) + V_{\mathrm{e}}(\mathbf{r})\sum_n F_n^k(\mathbf{r})U_n^k(\mathbf{r}) \\
&= E\sum_n F_n^k(\mathbf{r})U_n^k(\mathbf{r}) \qquad (3.5)
\end{aligned}$$

where in the first three terms of the left-hand side we have worked out the Laplacian and where n is understood to run over all included bands. The third term can be simplified as follows [52]:

$$\begin{aligned}
&\sum_n F_n^k(\mathbf{r})\left[\frac{-\hbar^2}{2m_e}\nabla^2 + V_{\mathrm{c}}(\mathbf{r})\right]U_n^k(\mathbf{r}) \\
&\qquad = \sum_n F_n^k(\mathbf{r})\left[\frac{-\hbar^2}{2m_e}\nabla^2 + \sum_l \theta_l(\mathbf{r})V_c^l(\mathbf{r})\right]U_n^k(\mathbf{r}) \\
&\qquad = \sum_n F_n^k(\mathbf{r})\sum_l \theta_l(\mathbf{r})\left[\frac{-\hbar^2}{2m_e}\nabla^2 + V_c^l(\mathbf{r})\right]U_n^k(\mathbf{r}) \\
&\qquad = \sum_n F_n^k(\mathbf{r})\sum_l \theta_l(\mathbf{r})\hat{H}^l(\mathbf{r})U_n^k(\mathbf{r}) \qquad (3.6)
\end{aligned}$$

where $\hat{H}^l$ is the bulk Hamiltonian of layer l. We now use the following completeness and orthonormality relations of the chosen basis functions in Eq. (3.5):

$$\begin{aligned}
&\sum_n U_n^{k*}(\mathbf{r}')U_n^k(\mathbf{r}) = \delta(\mathbf{r}-\mathbf{r}') \\
&\int_{\mathbb{R}^3} d\mathbf{r}\, U_m^{k*}(\mathbf{r})U_n^k(\mathbf{r}) = \delta_{nm} \qquad (3.7)
\end{aligned}$$

and equate the corresponding factors of U_n^k on both sides of the equation. The final system of N EF equations is then obtained as [52, 54]:

$$\frac{-\hbar^2}{2m_e}\nabla^2 F_n^k(\mathbf{r}) - \frac{i\hbar}{m_e}\sum_m \mathbf{p}_{nm}^k \cdot \nabla F_m^k(\mathbf{r}) + \sum_m H_{nm}^k(\mathbf{r}) F_m^k(\mathbf{r}) + V_e(\mathbf{r}) F_n^k(\mathbf{r}) = E F_n^k(\mathbf{r}) \tag{3.8}$$

with H_{nm}^k the interband Hamiltonian matrix elements and $\mathbf{p}_{nm}^k = (p_{nm}^{k,x}, p_{nm}^{k,y}, p_{nm}^{k,z})$ the interband momentum matrix elements referred to material layer k. The interband matrix elements provide the coupling between the different equations of the system: they couple band n, corresponding with F_n, to band m, corresponding to F_m. Hence, it is through the interband matrix elements that BTBT is described. The full definition of the interband Hamiltonian matrix elements and the interband momentum matrix elements is as follows:

$$\begin{aligned} H_{nm}^k(\mathbf{r}) &= \sum_l \theta_l(\mathbf{r}) \int_{V_{\text{uc}}} U_n^{k*}(\mathbf{r}') \hat{H}^l(\mathbf{r}') U_m^k(\mathbf{r}') d\mathbf{r}' \\ &= \sum_l \theta_l(\mathbf{r}) \sum_{i,j} \left[S^{l\to k}\right]_{in}^{\dagger} H_{ij}^l S_{mj}^{l\to k} \\ &= \sum_l \theta_l(\mathbf{r}) \sum_i \left[S^{l\to k}\right]_{in}^{\dagger} E_i^l S_{mi}^{l\to k} \end{aligned} \tag{3.9}$$

$$\begin{aligned} \mathbf{p}_{nm}^k &= -i\hbar \int_{V_{\text{uc}}} U_n^{k*}(\mathbf{r}') \nabla U_m^k(\mathbf{r}') d\mathbf{r}' \\ &= \sum_l \theta_l(\mathbf{r}) \sum_{i,j} \left[S^{l\to k}\right]_{in}^{\dagger} \mathbf{p}_{ij}^l S_{mj}^{l\to k} \end{aligned} \tag{3.10}$$

where V_{uc} is the volume of a unit cell, i is an index running over all bands, E_i^l are the bulk band edge eigenenergies in layer l and $\mathbf{p}_{ij}^l$ are the bulk interband momentum matrix elements of layer l. We have applied the heterostructure transformation of Van de Put *et al.* to obtain Eqs. (3.9) and (3.10) [52], which have a link to experimental data as explained at the end of this section. $S^{l\to k}$ is a unitary transformation of the basis functions from layer l to layer k based on a common eigenvalue decomposition of the momentum matrices [52]:

$$\mathbf{P}^k = \overline{Q^k}\,\text{diag}(\hbar\mathbf{G}) Q^{k\intercal} \tag{3.11}$$

where $\mathbf{P}^k = (P^{k,x}, P^{k,y}, P^{k,z})$ is a vector of the interband momentum matrices of material k, $\mathbf{G}$ is a reciprocal lattice vector and Q^k is the matrix containing the common eigenvectors of $P^{k,x}$, $P^{k,y}$ and $P^{k,z}$ with $\overline{Q^k}$ its element-wise complex conjugate. The transformation is then constructed as:

$$S^{l\to k} = Q^{k\dagger} Q^l \tag{3.12}$$

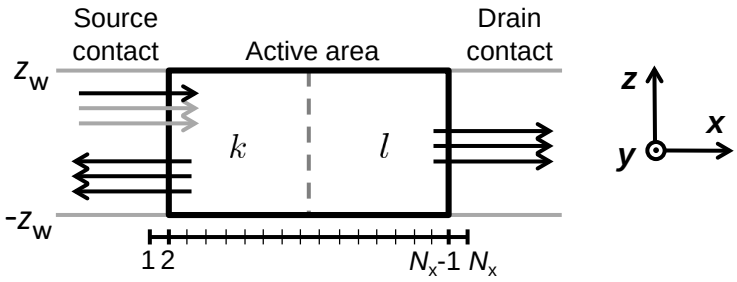

Figure 3.1: Schematic of a heterostructure device consisting of material layers k and l, where the heterojunction is indicated by the dashed line. The arrows represent subband modes. One mode is injected at the source contact, resulting in transmitted modes at the drain and reflected modes at the source. Also indicated is the discretization of the transport direction x and the boundaries z_w and $-z_w$ of the confined z-direction. The y-direction is translationally invariant.

with Q^l analogously obtained from a common eigenvalue decomposition of $\mathbf{P}^l$. The transformation matrix elements, used in Eqs. (3.9) and (3.10), are defined as:

$$S_{ab}^{l\to k} = \langle U_a^k | S^{l\to k} | U_b^k \rangle. \tag{3.13}$$

We can now see one of the key advantages of the EF formalism: in the system of Eq. (3.8), the highly oscillating basis functions U_n^k have been compounded into the matrix elements H_{nm}^k and $\mathbf{p}_{nm}^k$. H_{nm}^k can be constructed based on the bulk band edge energies E_i^l, which can be determined in optical experiments and are available in literature [139]. $\mathbf{p}_{nm}^k$ is composed of the bulk coupling parameters $\mathbf{p}_{ij}^l$ used in $\mathbf{k}\cdot\mathbf{p}$-theory. These can be obtained by fitting the $\mathbf{k}\cdot\mathbf{p}$ band structure to known experimental effective masses and energy gaps.

To limit the computational burden, we only consider the potential and material profiles to vary in two directions: the transport direction x and the confined direction z (see Fig. 3.1 for a general device schematic). The y-direction is considered infinitely long and translationally invariant. The EFs therefore have the following form:

$$F_n^k(\mathbf{r}) = e^{ik_y y} F_n^k(x,z) \tag{3.14}$$

with k_y the wave number in the y-direction. The EF system of Eq. (3.8) is then reduced to:

$$\begin{aligned}
&\frac{-\hbar^2}{2m_e}\frac{\partial^2 F_n^k(x,z)}{\partial x^2} + k_y^2\frac{\hbar^2}{2m_e}F_n^k(x,z) - \frac{\hbar^2}{2m_e}\frac{\partial^2 F_n^k(x,z)}{\partial z^2} \\
&- \frac{i\hbar}{m_e}\sum_m p_{nm}^{k,x}(x,z)\frac{\partial F_m^k(x,z)}{\partial x} + k_y\frac{\hbar}{m_e}\sum_m p_{nm}^{k,y}(x,z)F_m^k(x,z) \\
&- \frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x,z)\frac{\partial F_m^k(x,z)}{\partial z} + \sum_m H_{nm}^k(x,z)F_m^k(x,z) \\
&+ V_{\mathrm{e}}(x,z)F_n^k(x,z) = EF_n^k(x,z)
\end{aligned} \tag{3.15}$$

which is again a system of N equations. This EF system forms the basis of our QM simulation approach. It will continue to play a central role in subsequent chapters, where it will be modified for band structure models with a specific number of bands.

Before we can solve the system in Eq. (3.15) for the EFs, we first need to define appropriate boundary conditions.

3.3 Boundary conditions

3.3.1 Confined boundary conditions

In the confined z-direction, we assume Dirichlet hard wall boundary conditions for the wave function at the boundaries defined by $-z_{\mathrm{w}}$ and z_{w} in Fig. 3.1, i.e. $\psi(\mathbf{r}) = 0$ at $|z| = z_{\mathrm{w}}$. This means the wave function does not penetrate into the oxide, an approximation that becomes less valid for very confined structures.

3.3.2 EF form in the contacts

At the source and drain contacts, we assume the device to be open to allow for a current to flow. These open boundary conditions are implemented using QTBM. As introduced in Section 2.5.2, QTBM entails the fundamental assumption that in the source and drain contacts the electrostatic potential is constant in the transport direction [60]. This is justified by the fact that contact leads are generally metallic and do not possess variation or structure in the longitudinal direction. On the other hand, variation of the electrostatic potential in the

confined direction is allowed. Based on these assumptions, the EFs take a specific form in the source and drain contacts: a plane wave in the transport direction x, modulated by an envelope χ in the confined direction z, such that:

$$F_n(x,z) = \sum_\alpha c_\alpha e^{ik_{x,\alpha}x}\chi_{n\alpha}(z) \tag{3.16}$$

for a given energy E and k_y. Because of confinement in z, the EFs are a sum of subband modes. The index α counts all subband modes at a given E and k_y, with c_α as the corresponding complex coefficients. The subband modes in Eq. (3.16) can be sorted as incoming and outgoing based on the sign of the associated probability current, as we will show in Section 3.3.4. For the source contact, we define:

$$F_n(x,z) = \sum_\gamma I_\gamma e^{ik_{x,\gamma}x}\chi_{n\gamma}(z) + \sum_{\gamma'} r_{\gamma'} e^{ik_{x,\gamma'}x}\chi_{n\gamma'}(z) \tag{3.17}$$

with I_γ and $r_{\gamma'}$ denoting the coefficients of respectively the incoming and the outgoing modes. At the drain side, we define:

$$F_n(x,z) = \sum_\delta t_\delta e^{ik_{x,\delta}x}\chi_{n\delta}(z) + \sum_{\delta'} U_{\delta'} e^{ik_{x,\delta'}x}\chi_{n\delta'}(z) \tag{3.18}$$

where t_δ are the coefficients of the outgoing modes and $U_{\delta'}$ are the coefficients of the incoming modes. The forms in Eqs. (3.17) and (3.18) can serve as boundary conditions to the EF system in Eq. (3.15). I_γ and $U_{\delta'}$ are thereby used to inject modes into the device, while $r_{\gamma'}$ and t_δ are determined by the solution of the EF system. We illustrate this point in Section 3.4.2. To construct the forms in Eqs. (3.17) and (3.18), however, we first need to determine the different $k_{x,\alpha}$ and $\chi_{n\alpha}(z)$ at each contact and sort them into incoming and outgoing modes.

3.3.3 The contact eigenvalue problem

We can obtain the required $k_{x,\alpha}$ and $\chi_{n\alpha}(z)$ by solving an eigenvalue problem at each contact. To obtain this eigenvalue problem, we first insert Eq. (3.16)

into Eq. (3.15):

$$
\begin{aligned}
&k_{x,\alpha}^2 \frac{\hbar^2}{2m_e}\chi_{n\alpha}(z) + k_y^2 \frac{\hbar^2}{2m_e}\chi_{n\alpha}(z) - \frac{\hbar^2}{2m_e}\frac{d^2\chi_{n\alpha}(z)}{dz^2} \\
&+ k_{x,\alpha}\frac{\hbar}{m_e}\sum_m p_{nm}^x(x_\mathrm{c})\chi_{m\alpha}(z) + k_y \frac{\hbar}{m_e}\sum_m p_{nm}^y(x_\mathrm{c})\chi_{m\alpha}(z) \\
&- \frac{i\hbar}{m_e}\sum_m p_{nm}^z(x_\mathrm{c})\frac{d\chi_{m\alpha}(z)}{dz} + \sum_m H_{nm}(x_\mathrm{c})\chi_{m\alpha}(z) \\
&+ V_\mathrm{e}(x_\mathrm{c}, z)\chi_{n\alpha}(z) = E\chi_{n\alpha}(z),
\end{aligned}
\tag{3.19}
$$

which is a set of $N \cdot N_\alpha$ equations with N_α the number of subbands and x_c the x-location of the contact. Analogous to the derivation of Liu *et al.* for a 1D system [140], we formulate Eq. (3.19) as an eigenvalue problem:

$$
\mathbf{H}_\mathrm{c}(x_\mathrm{c}, z, k_{x,\alpha}, k_y)\boldsymbol{\chi}_\alpha(z) = E\boldsymbol{\chi}_\alpha(z). \tag{3.20}
$$

with $\boldsymbol{\chi}_\alpha(z)$ a vector containing for each band the confined envelopes:

$$
\boldsymbol{\chi}_\alpha(z) = \left[\chi_{1\alpha}(z)\chi_{2\alpha}(z)\chi_{3\alpha}(z)\cdots\chi_{N\alpha}(z)\right]^\mathsf{T} \tag{3.21}
$$

The contact Hamiltonian in Eq. (3.20) can be split up into terms which are quadratic, linear and constant in $k_{x,\alpha}$:

$$
\mathbf{H}_\mathrm{c}(x_\mathrm{c}, z, k_{x,\alpha}, k_y) = \mathbf{H}_\mathrm{c}^{(2)}k_{x,\alpha}^2 + \mathbf{H}_\mathrm{c}^{(1)}(x_\mathrm{c})k_{x,\alpha} + \mathbf{H}_\mathrm{c}^{(0)}(x_\mathrm{c}, z, k_y). \tag{3.22}
$$

where $\mathbf{H}_\mathrm{c}^{(2)}$, $\mathbf{H}_\mathrm{c}^{(1)}$ and $\mathbf{H}_\mathrm{c}^{(0)}$ are matrices that collect the respective coefficients. Finally, to obtain the $\boldsymbol{\chi}_\alpha$ and $k_{x,\alpha}$ needed to construct the EFs in the contact according to the form in Eq. (3.16), we rewrite the eigenvalue problem of Eq. (3.20) as:

$$
\begin{bmatrix} \mathbf{0} & \mathbf{1} \\ -(\mathbf{H}_\mathrm{c}^{(2)})^{-1}[\mathbf{H}_\mathrm{c}^{(0)} - E] & -(\mathbf{H}_\mathrm{c}^{(2)})^{-1}\mathbf{H}_\mathrm{c}^{(1)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\chi}_\alpha(z) \\ k_{x,\alpha}\boldsymbol{\chi}_\alpha(z) \end{bmatrix} = k_{x,\alpha} \begin{bmatrix} \boldsymbol{\chi}_\alpha(z) \\ k_{x,\alpha}\boldsymbol{\chi}_\alpha(z) \end{bmatrix}. \tag{3.23}
$$

For a given E and k_y, we can now solve Eq. (3.23) and obtain the $k_{x,\alpha}$ as eigenvalues, while the corresponding $\boldsymbol{\chi}_\alpha$ can be extracted from the eigenvectors. We do this for both the source and drain contact, assuming Dirichlet boundary conditions in the z-direction: $\boldsymbol{\chi}(z) = \mathbf{0}$ if $z = z_w$ or $z = -z_w$.

3.3.4 Sorting the contact modes

After having obtained a set of $k_{x,\alpha}$ and $\boldsymbol{\chi}_\alpha$ corresponding to the subband modes at each contact, the next step is to sort them into incoming and outgoing modes

at each contact. We do this based on the sign of their probability current in the x-direction, J_p^x. An expression for J_p^x can be derived from the formula for the probability current of a set of envelope functions [141]:

$$J_p^x = \int dz \frac{1}{m_e} \mathrm{Re}\left\{ \sum_n F_n^*(x,z) \left(-i\hbar \frac{\partial F_n(x,z)}{\partial x} \right) + \sum_{n,m} F_n^*(x,z) p_{nm}^x F_m(x,z) \right\} \tag{3.24}$$

where the integral in the z-direction runs over the contact. If we assume $k_{x,\alpha}$ to be a general, complex number $q_\alpha - \kappa_\alpha i$, then Eq. (3.16) results in the following form for the EFs in the contacts:

$$F_n(x,z) = \sum_\alpha c_\alpha \chi_{n\alpha}(z) e^{iq_\alpha x} e^{\kappa_\alpha x} \tag{3.25}$$

Inserting this expression into Eq. (3.24), we obtain for a single mode:

$$J_p^x = \frac{e^{2\kappa_\alpha x}|c_\alpha|^2}{m_e} \int dz \,\mathrm{Re}\left\{ \sum_n |\chi_{n\alpha}(z)|^2 \left(\hbar q_\alpha - i\hbar\kappa_\alpha\right) + \sum_{n,m} \chi_{n\alpha}^*(z) p_{nm}^x \chi_{m\alpha}(z) \right\} \tag{3.26}$$

where the integral over z runs over the width of the contact. Taking the real part of the first term, we obtain:

$$J_p^x = \frac{e^{2\kappa_\alpha x}|c_\alpha|^2}{m_e} \int dz \left[\hbar q_\alpha \sum_n |\chi_{n\alpha}(z)|^2 + \mathrm{Re}\left\{ \sum_{n,m} \chi_{n\alpha}^*(z) p_{nm}^x \chi_{m\alpha}(z) \right\} \right] \tag{3.27}$$

or written more compactly in a vector notation:

$$J_p^x = \frac{e^{2\kappa_\alpha x}|c_\alpha|^2}{m_e} \int dz \left[\hbar q_\alpha \left\| \boldsymbol{\chi_\alpha}(z) \right\|^2 + \mathrm{Re}\left\{ \boldsymbol{\chi_\alpha^*}(z) \mathbf{P} \boldsymbol{\chi_\alpha}(z) \right\} \right] \tag{3.28}$$

We can use this expression to sort all obtained subband modes. Since the prefactor of Eq. (3.28) is always positive, we determine the direction of the probability current from the sign of

$$\int dz \left[\hbar q_\alpha \left\| \boldsymbol{\chi_\alpha}(z) \right\|^2 + \mathrm{Re}\left\{ \boldsymbol{\chi_\alpha^*}(z) \mathbf{P} \boldsymbol{\chi_\alpha}(z) \right\} \right] \tag{3.29}$$

Note that the direction of the probability current of a given mode does not necessarily correspond to the sign of the corresponding k_x-value. Through the momentum matrix $\mathbf{P}$, the sign of J_p^x is also determined by the local band curvature. In regions with a negative curvature, such as the valence band for instance, a right going mode generally corresponds to a negative k_x-value (see Fig. 3.2).

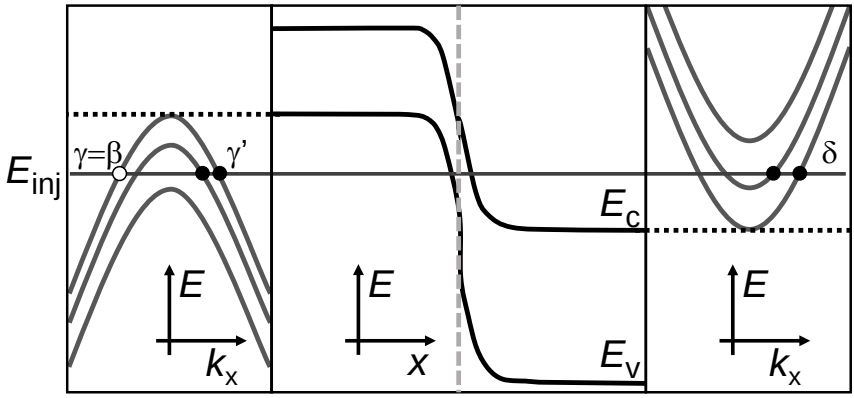

Figure 3.2: Schematic depiction of the injection of a single mode (open circle) into a heterojunction diode at an injection energy E_{inj}. The probability current is transmitted to two modes at the other contact and reflected to two modes at the injection contact (closed circles). The Greek symbols correspond to the respective mode indices. Notice that the mode β has an ingoing probability current, despite corresponding to a negative k_x-value. Both valence and conduction band subbands have been depicted as simplified parabolas for clarity. The dashed line indicates the location of the heterojunction.

Using Eq. (3.29) to sort the modes, we can construct the EFs in the contacts using Eqs. (3.17) and (3.18). These form the boundary conditions to the EF system of Eq. (3.15). The next step is then to develop a numerical implementation for this system and its boundary conditions.

3.4 Numerical discretization

3.4.1 EF system discretization

To solve the EF system of Eq. (3.15) numerically, we discretize the x-and z-direction in N_x and N_z points respectively. With N the number of bands (EFs), we can then define a discretized EF vector of dimensions ($N_x \cdot N_z \cdot N \times 1$), containing the values of all EFs at all mesh points:

$$\begin{aligned}\mathbf{F} &= \left[\mathbf{F}_{1,z}\mathbf{F}_{2,z}\cdots\mathbf{F}_{N_x,z}\right]^\intercal \\ &= \left[\left[\mathbf{F}_{1,1}\mathbf{F}_{1,2}\cdots\mathbf{F}_{1,N_z}\right]\left[\mathbf{F}_{2,1}\mathbf{F}_{2,2}\cdots\mathbf{F}_{2,N_z}\right]\cdots\left[\mathbf{F}_{N_x,1}\mathbf{F}_{N_x,2}\cdots\mathbf{F}_{N_x,N_z}\right]\right]^\intercal,\end{aligned} \tag{3.30}$$

with

$$\mathbf{F}_{i,j} = \left[F_1(x_i, z_j)F_2(x_i, z_j)\cdots F_N(x_i, z_j)\right]^\intercal . \tag{3.31}$$

The partial derivatives of the EFs are approximated with central FDs [61]. For the partial derivative in x at the point (x_i,z_j), this yields:

$$\left.\frac{\partial F(x,z)}{\partial x}\right|_{x_i,z_j} = \frac{F(x_{i+1}, z_j) - F(x_{i-1}, z_j)}{2\Delta x} \tag{3.32}$$

with Δx the discretization step size. A partial derivative can hence be represented by the multiplication of the discretized EF vector with a matrix:

$$\mathbf{D1}_x = \frac{1}{2\Delta x}\begin{bmatrix} 0 & 1 & & & & & \\ -1 & 0 & 1 & & & & \\ & -1 & 0 & 1 & & & \\ & & -1 & 0 & 1 & & \\ & & & & \ddots & & \\ & & & & & -1 & 0 \end{bmatrix} \tag{3.33}$$

The second order derivatives are obtained as:

$$\left.\frac{\partial^2 F(x,z)}{\partial x^2}\right|_{x_i,z_j} = \frac{F(x_{i+1}, z_j) - 2F(x_i, z_j) + F(x_{i-1}, z_j)}{\Delta x^2} \tag{3.34}$$

which can be represented by a matrix multiplication with

$$\mathbf{D2}_x = \frac{1}{\Delta x^2}\begin{bmatrix} -2 & 1 & & & & & \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & 1 & -2 & 1 & & \\ & & & & \ddots & & \\ & & & & & 1 & -2 \end{bmatrix} \tag{3.35}$$

The partial derivatives in z are analogous. Eqs. (3.33) and (3.35) allow us to write Eq. (3.15) as a discretized matrix equation:

$$\mathbf{KF} = E\mathbf{F} \tag{3.36}$$

where $\mathbf{K}$ is the discretized EF system matrix, with dimensions ($N_x \cdot N_z \cdot N \times N_x \cdot N_z \cdot N$). Instructions on how to construct $\mathbf{K}$ with $\mathbf{D1}_x$ and $\mathbf{D2}_x$ can be found in Appendix A. Note that in Chapter 5, the FD discretization in z is replaced with a spectral method for higher band models.

3.4.2 QTBM boundary condition discretization

Having discretized the EF system, the next step is to discretize the QTBM contact eigenvalue problem and incorporate the boundary conditions into the discretized EF system matrix $\mathbf{K}$. Analogous to the EF vector defined in Eq. (3.30), the contact envelopes $\chi_{n\alpha}$ are grouped and discretized as follows:

$$\boldsymbol{\chi}_{\alpha}^{S/D} = \left[\boldsymbol{\chi}_{\alpha,1}^{S/D}\boldsymbol{\chi}_{\alpha,2}^{S/D}\boldsymbol{\chi}_{\alpha,3}^{S/D}\cdots\boldsymbol{\chi}_{\alpha,N_z}^{S/D}\right]^{\mathsf{T}} \tag{3.37}$$

with

$$\boldsymbol{\chi}_{\alpha,j}^{S/D} = \left[\chi_{1\alpha}(z_j)\chi_{2\alpha}(z_j)\cdots\chi_{N\alpha}(z_j)\right] \tag{3.38}$$

where the superscript S (D) denotes a discretized vector at the source (drain) contact. The discretization of the contact eigenvalue problem in Eq. (3.23) then proceeds analogously to the EF system.

Next, we connect the solutions of the discretized contact eigenvalue problems with the discretized EF system matrix $\mathbf{K}$. To do this, we define the two edge points at each side of the mesh to be inside the contact region. These points are labeled 1 and 2 at the source and $N_x - 1$ and N_x at the drain (see Fig. 3.1). For these contact points, the EFs take the form specified in Eqs. (3.17) and (3.18). For the point at each contact with the lowest x-value, namely 1 and $N_x - 1$, we can write the following discretized matrix equations:

$$\mathbf{F}_{1,z} = \mathbf{S}_{11}\mathbf{I} + \mathbf{S}_{12}\mathbf{r} \tag{3.39}$$

$$\mathbf{F}_{N_x-1,z} = \mathbf{D}_{11}\mathbf{t} + \mathbf{D}_{12}\mathbf{U} \tag{3.40}$$

where the vector $\mathbf{I}$ ($\mathbf{U}$) contains the coefficients of the incoming modes at the source (drain) and $\mathbf{r}$ ($\mathbf{t}$) contains the coefficients of the outgoing modes at the source (drain). The $\mathbf{S}$ and $\mathbf{D}$-matrices are constructed from the $\boldsymbol{\chi}_{\alpha}^{S/D}$-vectors and $k_{x,\alpha}^{S/D}$-values of the source and drain respectively. For the source contact:

$$\mathbf{S}_{11} = \left[\boldsymbol{\chi}_{1}^{S}, \boldsymbol{\chi}_{2}^{S}, \ldots, \boldsymbol{\chi}_{N_\mathrm{R}}^{S}\right] \tag{3.41}$$

$$\mathbf{S}_{12} = \left[\boldsymbol{\chi}_{N_\mathrm{R}+1}^{S}, \boldsymbol{\chi}_{N_\mathrm{R}+2}^{S}, \ldots, \boldsymbol{\chi}_{N_\mathrm{R}+N_\mathrm{L}}^{S}\right] \tag{3.42}$$

with N_R (N_L) the number of right (left) propagating modes. The $\mathbf{D}$-matrices are analogously constructed with the $\boldsymbol{\chi}^{D}$-vectors. For mesh points 2 and N_x, which are one discretized x-value higher than 1 and $N_x - 1$ respectively, we write matrix equations similar to Eqs. (3.39) and (3.40):

$$\mathbf{F}_{2,z} = \mathbf{S}_{21}\mathbf{I} + \mathbf{S}_{22}\mathbf{r} \tag{3.43}$$

$$\mathbf{F}_{N_x,z} = \mathbf{D}_{21}\mathbf{t} + \mathbf{D}_{22}\mathbf{U} \tag{3.44}$$

except now the **S** and **D**-matrices are shifted over one discretization step Δx:

$$\mathbf{S}_{21} = \left[e^{ik^S_{x,1}\Delta x}\boldsymbol{\chi}^S_1, e^{ik^S_{x,2}\Delta x}\boldsymbol{\chi}^S_2, \ldots, e^{ik^S_{x,N_\mathrm{R}}\Delta x}\boldsymbol{\chi}^S_{N_\mathrm{R}} \right] \tag{3.45}$$

$$\mathbf{S}_{22} = \Big[e^{ik^S_{x,N_\mathrm{R}+1}\Delta x}\boldsymbol{\chi}^S_{N_\mathrm{R}+1},\ e^{ik^S_{x,N_\mathrm{R}+2}\Delta x}\boldsymbol{\chi}^S_{N_\mathrm{R}+2}, \ldots, e^{ik^S_{x,N_\mathrm{R}+N_\mathrm{L}}\Delta x}\boldsymbol{\chi}^S_{N_\mathrm{R}+N_\mathrm{L}} \Big] \tag{3.46}$$

From Eqs. (3.39) and (3.40) and Eqs. (3.43) and (3.44), $\mathbf{r}$ and $\mathbf{t}$ can be eliminated, resulting in the discretized QTBM boundary conditions:

$$\mathbf{F}_{1,z} - \mathbf{S}_{12}\mathbf{S}^{-1}_{22}\mathbf{F}_{2,z} = \mathbf{S}_{11}\mathbf{I} - \mathbf{S}_{12}\mathbf{S}^{-1}_{22}\mathbf{S}_{21}\mathbf{I} \tag{3.47}$$

$$-\mathbf{D}_{21}\mathbf{D}^{-1}_{11}\mathbf{F}_{N_x-1,z} + \mathbf{F}_{N_x,z} = \mathbf{D}_{22}\mathbf{U} - \mathbf{D}_{21}\mathbf{D}^{-1}_{11}\mathbf{D}_{12}\mathbf{U}. \tag{3.48}$$

The coefficient vectors $\mathbf{I}$ and $\mathbf{U}$ are set depending on which mode is to be injected. For example, if we want to inject a mode β at the source side (see Fig. 3.2), we set the corresponding element I_β of $\mathbf{I}$ to 1 and all other elements $I_{\beta\neq\gamma}$ and $U_{\delta'}$ to 0. To inject modes at the drain, we do the same with $\mathbf{U}$. We will use the injection of individual modes to determine transmission probabilities that are needed for the calculation of the device current.

3.4.3 Full system

We now combine the discretized EF system with the discretized boundary conditions. The rows of $\mathbf{K}$ corresponding to the source and drain contacts are replaced by the left-hand side of Eqs. (3.47) and (3.48) respectively, resulting in the new matrix $\mathbf{K}_\mathrm{bound}$, while the right-hand sides form an additional source term $\mathbf{R}$. Eq. (3.36) then becomes a linear system:

$$\begin{aligned}(\mathbf{K}_\mathrm{bound} - E\mathbf{I}_\mathrm{tot})\mathbf{F} &= \mathbf{R} \\ \mathbf{K}_\mathrm{tot}\mathbf{F} &= \mathbf{R}\end{aligned} \tag{3.49}$$

with $\mathbf{I}_\mathrm{tot}$ the $N_x \cdot N_z \cdot N \times N_x \cdot N_z \cdot N$ unity matrix. Eq. (3.49) can be solved for $\mathbf{F}$. Note that for the injection of a different mode at the same E and k_y, only the right-hand side changes, since it contains $\mathbf{I}$ and $\mathbf{U}$. We therefore apply a sparse LU factorization to $\mathbf{K}_\mathrm{tot}$, such that we can reuse this factorization for all modes. Eq. (3.49) becomes:

$$\begin{aligned}\mathbf{LUF} &= \mathbf{PR} \\ \mathbf{F} &= \mathbf{PR}\backslash\mathbf{L}\backslash\mathbf{U}\end{aligned} \tag{3.50}$$

where $\mathbf{L}$,$\mathbf{U}$ and $\mathbf{P}$ are respectively a lower triangular, upper triangular and permutation matrix, all independent of the right-hand side $\mathbf{R}$. The solution of Eq. (3.50) results in the desired EF vector, which contains the value of all EFs at all mesh points (see Eq. (3.30)). We can now use $\mathbf{F}$ to calculate the device current.

3.5 Current calculation

The calculation of the actual electrical current in the device is based on the Landauer formalism for ballistic transport: we first determine transmission probabilities of the individual subband modes and then weight them with the appropriate occupation probabilities. These contributions are finally summed and integrated over all degrees of freedom.

3.5.1 Transmission probability

The transmission probability of a single mode can be found from the ratio of transmitted and injected probability current when only that mode is injected. According to Eq. (3.28), the injected probability current for a given mode β at the source is:

$$J_p^{x,\mathrm{IN}} = \int dz \left[\frac{\hbar q_\beta |I_\beta|^2}{m_e} \left\| \boldsymbol{\chi}_\beta(z) \right\|^2 + \frac{|I_\beta|^2}{m_e} \mathrm{Re}\left\{ \boldsymbol{\chi}_\beta^*(z) \mathbf{P} \boldsymbol{\chi}_\beta(z) \right\} \right]. \tag{3.51}$$

The transmitted probability current is similarly obtained from the transmitted subband modes at the drain contact:

$$J_p^{x,\mathrm{OUT}} = \sum_\delta \left[\int dz \left[\frac{\hbar q_\delta |t_\delta|^2}{m_e} \left\| \boldsymbol{\chi}_\delta(z) \right\|^2 + \frac{|t_\delta|^2}{m_e} \mathrm{Re}\left\{ \boldsymbol{\chi}_\delta^*(z) \mathbf{P} \boldsymbol{\chi}_\delta(z) \right\} \right] \right] \tag{3.52}$$

where the index δ runs over all outgoing modes at the drain. When injecting at the source side, $\mathbf{U}$ is zero in Eq. (3.44), such that we can obtain the coefficients t_δ from a decomposition of the calculated EFs at the drain:

$$\mathbf{t} = (\mathbf{D}_{21})^{-1} \mathbf{F}_{N_x,z} \tag{3.53}$$

where $\mathbf{F}_{N_x,z}$ is extracted from the solution vector $\mathbf{F}$ of Eq. (3.49). For the injected mode β at the given E and k_y, the transmission probability is then given by:

$$T_\beta(E, k_y) = \frac{J_p^{x,\mathrm{OUT}}}{J_p^{x,\mathrm{IN}}} \tag{3.54}$$

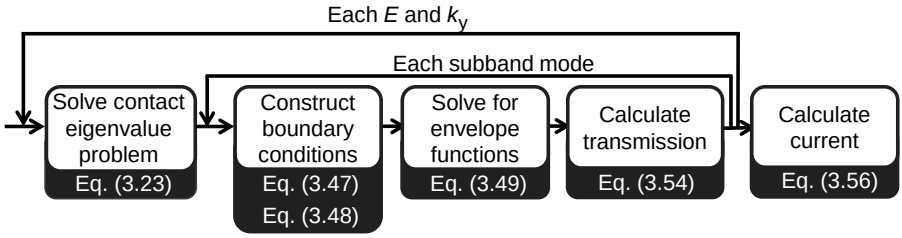

Figure 3.3: Flowchart of the formalism outlined in this chapter, along with the key equations for each step.

3.5.2 Current formula

To calculate the full current, we have to calculate the transmission probability for each ingoing subband mode as in the previous section, and do this for each E and k_y. This is only required for one of the contacts, since the transmission probability at a given E and k_y is independent of the injecting contact because of time reversal symmetry. We then obtain the total current density formula from the Tsu-Esaki formula for ballistic transport [50]:

$$I_{DS} = \frac{2q}{h} \int_E \int_{k_y} \int_{k_z} T(E, k_y, k_z)(f_S(E) - f_D(E)) \frac{dk_y}{2\pi} \frac{dk_z}{2\pi} dE, \tag{3.55}$$

where $f_{\mathrm{S}}(E)$ and $f_{\mathrm{D}}(E)$ are the Fermi-Dirac distributions respectively in the source and drain contact. Because of confinement in z, translational invariance is lost in that direction. The integral over k_z is therefore replaced with a discrete sum over the subband transmission probabilities:

$$I_{\mathrm{DS}} = \frac{2q}{h} \int_E \int_{k_y} \sum_{\gamma} T_\gamma(E, k_y)(f_{\mathrm{S}}(E) - f_{\mathrm{D}}(E)) \frac{dk_y}{2\pi} dE, \tag{3.56}$$

where the index γ runs over all ingoing subband modes at a given E and k_y. Note that I_{DS} is in units of A/m.

3.6 Conclusion

In this chapter, we presented a quantum mechanical formalism aimed at simulating heterostructure TFETs in a computationally efficient way. Based on these requirements, we opted for a multi-band 2D EF formalism, combined with the QTBM. We outlined the general formalism, showing how the EF system with its boundary conditions is constructed and discretized and how transmission

probabilities and current density can be obtained from its solutions. The full procedure is summarized in Fig. 3.3.

We developed the formalism for an arbitrary number of bands. For a given application, however, it is necessary to choose a specific basis set. Each choice of basis set comes with its own specific issues, often requiring a modification of the basic formalism outlined in this chapter. In the following chapters, we will therefore discuss different basis set options, starting with the most simple model and moving towards more complex higher band models. We will discuss the issues that come with the implementation of each basis set, develop solutions and apply the implemented models to simulate and optimize appropriate TFET configurations.

The formalism outlined in this chapter has been published in the *Journal of Applied Physics* [142].

Chapter 4

Two-band model

In the previous chapter, we developed a QM simulation approach for BTBT in devices with a heterostructure. The employed band structure model was based on an EF formalism valid for an arbitrary number of bands. To put our approach to use in simulating the performance of actual heterostructure TFET configurations, we now implement this model for a specific number of bands. In this chapter, we start with the simplest band model that allows for the modeling of BTBT: a two-band model, and use it to study the interplay between SIQC and gate control in heterostructure pointTFETs. Because of the small number of bands, the model is limited in accuracy and can only reliably simulate specially-designed pointTFETs, in which the tunneling is oriented predominantly parallel to the gate dielectric. In the next chapters, we will increase the number of bands to allow for the simulation of other TFET configurations, such as the lineTFET.

The chapter is structured as follows. We first discuss the chosen two-band basis set, the corresponding coupling parameters and the resulting EF system (Section 4.1). Next, we indicate the issues that arise with such a limited basis set, namely the presence of spurious solutions in the band structure and the lack of coupling between basis functions in the directions orthogonal to transport (Section 4.2). We develop a strategy to handle the spurious solutions and alleviate the lack of coupling by introducing effective mass terms into the EF system for the orthogonal directions. We then calibrate the parameters of the model with experimental diodes (Section 4.3). Finally, we use our two-band implementation to simulate transmission spectra and currents in straddled and staggered bandgap heterostructure diodes and staggered bandgap pocketed pointTFETs. We find a counteracting effect between SIQC and gate control

for decreasing body thickness (Section 4.4). The chapter ends with conclusions (Section 4.5).

4.1 Two-band basis set

4.1.1 Choice of basis functions

In a two-band model, the wave function is expanded on a set of two basis functions. The EF expansion of Eq. (3.4) then reduces to

$$\psi(\mathbf{r}) = F_1^k(\mathbf{r})U_1^k(\mathbf{r}) + F_2^k(\mathbf{r})U_2^k(\mathbf{r}) \tag{4.1}$$

where we have labeled the bands with the subscripts 1 and 2. To define the basis functions, it is common in literature to start from a set of abstract basis states. In our set, we include one state associated with the valence band, denoted as $|X\rangle$, and one state associated with the conduction band, $|S\rangle$. These states are so named because $|X\rangle$ has directional symmetry like a p_x-orbital ("p-type symmetry") and $|S\rangle$ has isotropic symmetry like an s-orbital ("s-type symmetry"). The basis functions are then defined as the basis states projected onto the position basis:

$$\begin{aligned} U_1^k(\mathbf{r}) &= \langle \mathbf{r}|X\rangle \\ U_2^k(\mathbf{r}) &= \langle \mathbf{r}|S\rangle \end{aligned} \tag{4.2}$$

To construct the EF system of Eq. (3.15), we also need the associated Hamiltonian matrix elements H_{nm}^k and the momentum matrix elements $\mathbf{p}_{nm}^k$, which are defined in Eqs. (3.9) and (3.10). In a two-band model, the heterotransformation reduces to a unity transformation [52]. The bulk Hamiltonian matrix therefore reduces to a diagonal matrix of the band edge energies in each material. Based on Eq. (3.9), we can then write the two non-zero Hamiltonian matrix elements as:

$$\begin{aligned} H_{11}^k(x,z) &= \sum_l \theta_l(x,z)E_{\mathrm{v}}^l \\ H_{22}^k(x,z) &= \sum_l \theta_l(x,z)E_{\mathrm{c}}^l \end{aligned} \tag{4.3}$$

with E_{c}^l and E_{v}^l the band edge energies in material layer l of the conduction and valence band respectively. For the momentum matrix elements, the symmetry

properties of the states result in only one non-zero element in each material layer [143]:

$$\begin{aligned} p_{12}^k &= \langle S| \, p_x^k \, |X\rangle \\ &= -i\hbar \int_{V_{\mathrm{uc}}} U_2^{k*}(\mathbf{r}') \frac{\partial U_1^k(\mathbf{r}')}{\partial x} d\mathbf{r}' \end{aligned} \tag{4.4}$$

with $p_{21}^k = p_{12}^{k*}$ since the momentum operator is Hermitian. For the other matrix elements, we have:

$$\langle S| \, p_y^k \, |X\rangle = \langle S| \, p_z^k \, |X\rangle = 0 \tag{4.5}$$

Taking into account that the heterotransformation is a unity transformation, we can write Eq. (3.10) as:

$$p_{12}^k = p_{21}^{k*} = p^k = \sum_l \theta_l(x,z) p^l \tag{4.6}$$

where we have dropped the subscripts on the only non-zero momentum matrix element in each layer.

4.1.2 Two-band EF system

We now insert the chosen two-band expansion of Eq. (4.1) and the expressions for the matrix elements in Eqs. (4.3) and (4.6) into the 2D EF system for an infinite number of bands of Eq. (3.15). This results in the following two-band EF system:

$$\begin{aligned} &\frac{-\hbar^2}{2m_e} \frac{\partial^2}{\partial x^2} \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} + k_y^2 \frac{\hbar^2}{2m_e} \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} - \frac{\hbar^2}{2m_e} \frac{\partial^2}{\partial z^2} \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} \\ &- \frac{i\hbar}{m_e} \sum_l \theta_l(x,z) \begin{bmatrix} 0 & p^l \\ (p^l)^* & 0 \end{bmatrix} \frac{\partial}{\partial x} \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} + \sum_l \theta_l(x,z) \begin{bmatrix} E_{\mathrm{v}}^l & 0 \\ 0 & E_{\mathrm{c}}^l \end{bmatrix} \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} \\ &+ V_{\mathrm{e}}(x,z) \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} = E \begin{bmatrix} F_1^k(x,z) \\ F_2^k(x,z) \end{bmatrix} \end{aligned} \tag{4.7}$$

which is a system of two equations, here written as a matrix equation. The QTBM contact eigenvalue problem of Eq. (3.19) is similarly reduced to a set of

$2N_\alpha$ equations:

$$k_{x,\alpha}^2 \frac{\hbar^2}{2m_e}\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix} + k_y^2 \frac{\hbar^2}{2m_e}\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix} - \frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial z^2}\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix}$$
$$+ k_{x,\alpha}\frac{\hbar}{m_e}\begin{bmatrix}0 & p\\ p^* & 0\end{bmatrix}\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix} + \begin{bmatrix}E_\mathrm{v} & 0\\ 0 & E_\mathrm{c}\end{bmatrix}\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix}$$
$$+ V_\mathrm{e}(x_\mathrm{c}, z)\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix} = E\begin{bmatrix}\chi_{1\alpha}(z)\\ \chi_{2\alpha}(z)\end{bmatrix} \tag{4.8}$$

where we assume a single material over the contact. Note that Eq. (4.8) can also be used to plot the bulk band structure, by replacing the second derivative to z by k_z^2. This corresponds to assuming an infinitely long z-direction (see Fig. 4.1(a) for an example with InAs).

Following the approach outlined in the previous chapter for an arbitrary number of bands, the solutions of Eqs. (4.7) and (4.8) should enable the calculation of the tunneling probability, and eventually the current. The two-band model presents specific issues, however, which require a modified approach.

4.2 Model-specific issues

The two-band EF system and associated QTBM eigenvalue problem in Eqs. (4.7) and (4.8) have two important issues compared to the infinite band case. The first is the absence of coupling terms in the y- and z-direction. These terms have disappeared because the corresponding momentum matrix elements are zero (see Eq. (4.5)). The lack of coupling in the directions orthogonal to transport results in an upwards curvature of the valence bands in those directions (see Fig. 4.1(a)). The second issue is the appearance of spurious solutions in the band structure, which manifest themselves in the bulk band structure of Fig. 4.1(a) as an upwards curvature of the valence band outside of the first Brillouin zone in the transport direction. It is clear that we have to deal with these issues before we can attempt transport simulations based on Eqs. (4.7) and (4.8). In this section, we therefore further discuss the lack of coupling in Section 4.2.1 and subsequently modify the two-band EF system to correct it in Section 4.2.2. We then investigate the origin of the spurious solutions and develop a strategy to handle them in Section 4.2.3.

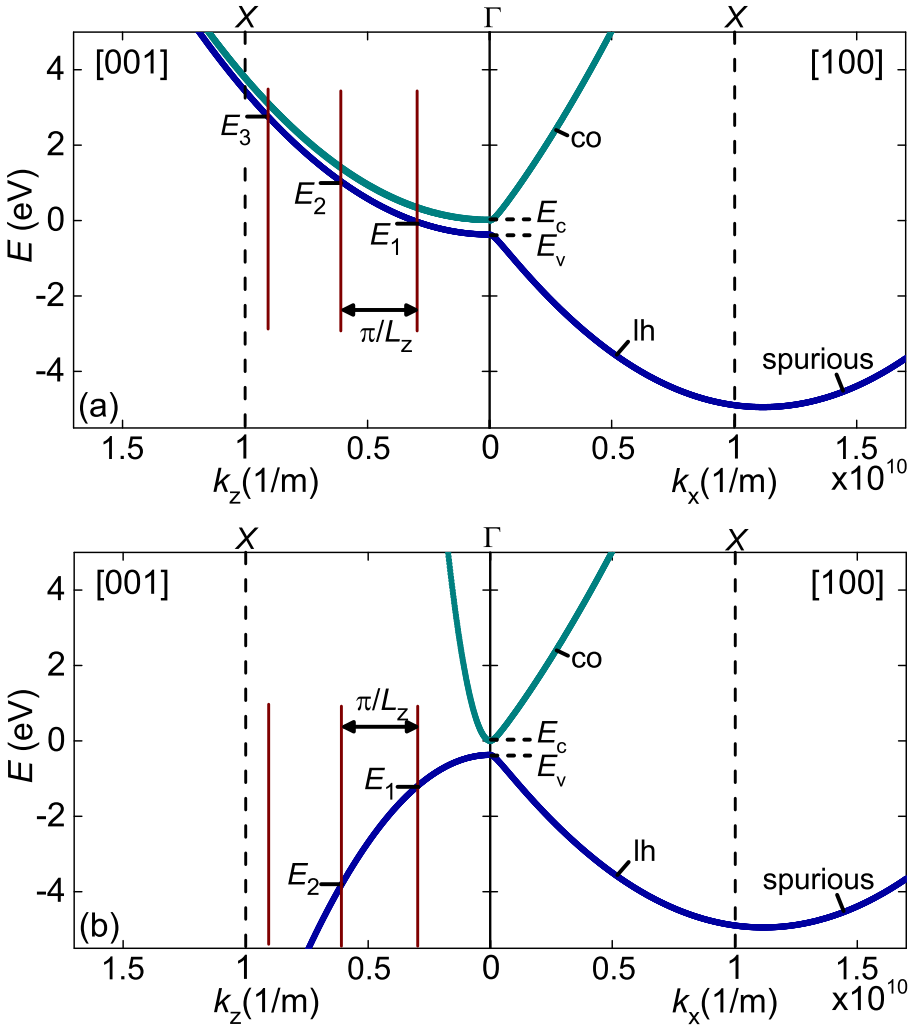

Figure 4.1: Two-band bulk band structure of InAs in the [100] transport direction (right) and one of the orthogonal directions (left), (a) before and (b) after the introduction of effective masses in the orthogonal directions. Three quantization slices (brown vertical lines) and subband energy levels (E_1, E_2 and E_3) are indicated for an L_z of 1 nm. The dashed vertical lines indicate the edges of the first Brillouin zone. The solutions are colored based on the dominant basis state, identified by the component with the largest absolute value in the corresponding vector $\boldsymbol{\chi}_\alpha$ (defined in Eq. (3.21)). Indicated are the conduction (co) and light hole (lh) band. The two-band parameters are listed in Appendix C.2.

4.2.1 Lack of orthogonal coupling

In the exact case of an infinite number of bands, the downwards curvature of the valence bands is assured through the coupling with conduction band basis functions. This coupling occurs through the terms linear in k and the terms with a first order derivative to x, y or z in the EF system of Eq. (3.15) and the contact eigenvalue problem of Eq. (3.19). In the infinite band case, the valence band consists of a combination of three basis states: $|X\rangle$, $|Y\rangle$ and $|Z\rangle$, so named because they exhibit directional p-type symmetry in respectively the x, y and z-direction. Because of this symmetry, each of these states couples with the first conduction band state $|S\rangle$ only in its corresponding direction and forms a strongly curved light hole band. In the other directions, the states only couple with conduction band states of higher energy. This coupling is weaker, and results in two heavy hole bands with a smaller curvature. In each direction we therefore end up with a light hole band, formed by the valence band state that couples with the lowest conduction band state in that direction, and two heavy hole bands, formed by the other two states. E.g. in the x-direction, $|X\rangle$ couples to $|S\rangle$ through the matrix element $\langle S|\, p_x\, |X\rangle$ and forms the light hole band, while $|Y\rangle$ and $|Z\rangle$ do not couple with $|S\rangle$ and form the heavy hole bands. The other directions are analogous. Note that in the presence of spin-orbit coupling, one of the heavy hole bands splits off (see Chapter 6).

In a two-band model, only the lowest conduction band state $|S\rangle$ and one valence band state $|X\rangle$ are included in the model. Hence, the only remaining coupling forms a light hole band in the transport direction x, with no heavy hole band present. In the other directions, y and z, the lack of coupling between $|S\rangle$ and $|X\rangle$ results in an upwards curving free electron band.

The consequences of the upwards curving valence band are an unphysical orthogonal DOS and a wrong ordering of valence band subbands in the presence of confinement. Fig. 4.1(a) shows that subband energy levels for a confined structure can be found from the bulk band structure by taking slices in the direction of the confinement. Because of the upwards curvature, the ordering of the subband energy levels of the valence band is opposite to what is physically expected, with energy levels ending up in the bandgap and conduction band. To enable transport calculations, we therefore first need to correct the orthogonal valence band curvature.

4.2.2 Modified EF system

We reintroduce the downwards curvature of the orthogonal valence bands by introducing two effective masses. These effective masses are a substitute for

the absent linear coupling terms and artificially incorporate the effect of bands that are not included in the model. We introduce the effective masses in the EF system of Eq. (4.7) in the terms quadratic in k_y and k_z. This results in:

$$\frac{-\hbar^2}{2m_e}\frac{\partial^2}{\partial x^2}\begin{bmatrix}F_1^k(x,z)\\F_2^k(x,z)\end{bmatrix} + k_y^2\frac{\hbar^2}{2}\begin{bmatrix}\frac{1}{m_{\mathrm{v}}^*(x,z)}F_1^k(x,z)\\\frac{1}{m_{\mathrm{c}}^*(x,z)}F_2^k(x,z)\end{bmatrix} - \frac{\hbar^2}{2}\frac{\partial^2}{\partial z^2}\begin{bmatrix}\frac{1}{m_{\mathrm{v}}^*(x,z)}F_1^k(x,z)\\\frac{1}{m_{\mathrm{c}}^*(x,z)}F_2^k(x,z)\end{bmatrix}$$

$$-\frac{i\hbar}{m_e}\sum_l \theta_l(x,z)\begin{bmatrix}0 & p^l\\(p^l)^* & 0\end{bmatrix}\frac{\partial}{\partial x}\begin{bmatrix}F_1^k(x,z)\\F_2^k(x,z)\end{bmatrix} + \sum_l \theta_l(\mathbf{r})\begin{bmatrix}E_{\mathrm{v}}^k & 0\\0 & E_{\mathrm{c}}^k\end{bmatrix}\begin{bmatrix}F_1^k(x,z)\\F_2^k(x,z)\end{bmatrix}$$

$$+V_{\mathrm{e}}(x,z)\begin{bmatrix}F_1^k(x,z)\\F_2^k(x,z)\end{bmatrix} = E\begin{bmatrix}F_1^k(x,z)\\F_2^k(x,z)\end{bmatrix} \tag{4.9}$$

with m_{c}^* and m_{v}^* the effective masses of conduction band and valence band respectively, which depend on position in the case of a heterostructure. The use of m_{c}^* and m_{v}^* provides a parabolic approximation of the actual non-parabolic DOS of the respective bands.

Since we can only introduce one effective mass for the valence band, a choice has to be made for m_{v}^* between the heavy and light hole mass. The effective masses determine the orthogonal DOS, so the question is whether the dominant DOS contribution to BTBT currents is determined by the light hole or the heavy hole band. As the coupling between conduction band and heavy hole band is weaker, the heavy hole DOS is much larger than that of the light hole band. Since one state has a heavy character in two directions and light character in the other direction, we assume the heavy hole band determines the orthogonal DOS. We verify this assumption with experimental data in Section 4.4.3 and with a higher band model in Section 5.5.3.

For the QTBM contact eigenvalue problem, we similarly introduce the effective masses in the contact eigenvalue problem of Eq. (4.8), yielding:

$$k_{x,\alpha}^2\frac{\hbar^2}{2m_e}\begin{bmatrix}\chi_{1,\alpha}(z)\\\chi_{2,\alpha}(z)\end{bmatrix} + k_y^2\frac{\hbar^2}{2}\begin{bmatrix}\frac{1}{m_{\mathrm{v}}^*}\chi_{1,\alpha}(z)\\\frac{1}{m_{\mathrm{c}}^*}\chi_{2,\alpha}(z)\end{bmatrix} - \frac{\hbar^2}{2}\frac{\partial^2}{\partial z^2}\begin{bmatrix}\frac{1}{m_{\mathrm{v}}^*}\chi_{1,\alpha}(z)\\\frac{1}{m_{\mathrm{c}}^*}\chi_{2,\alpha}(z)\end{bmatrix}$$

$$+k_{x,\alpha}\frac{\hbar}{m_e}\begin{bmatrix}0 & p\\p^* & 0\end{bmatrix}\begin{bmatrix}\chi_{1,\alpha}(z)\\\chi_{2,\alpha}(z)\end{bmatrix} + \begin{bmatrix}E_{\mathrm{v}} & 0\\0 & E_{\mathrm{c}}\end{bmatrix}\begin{bmatrix}\chi_{1,\alpha}(z)\\\chi_{2,\alpha}(z)\end{bmatrix}$$

$$+V_{\mathrm{e}}(x_{\mathrm{c}},z)\begin{bmatrix}\chi_{1,\alpha}(z)\\\chi_{2,\alpha}(z)\end{bmatrix} = E\begin{bmatrix}\chi_{1,\alpha}(z)\\\chi_{2,\alpha}(z)\end{bmatrix} \tag{4.10}$$

where we again assume a single material for the contact regions. If we now plot the bulk band structure based on Eq. (4.10) (see Fig. 4.1(b)), we see that the

valence band has a downwards curvature, with the ordering of the subband energy levels as physically expected.

Eqs. (4.9) and (4.10) are implemented with a finite difference discretization in both the x- and z-directions as described in Section 3.4 and Appendix A. We can now plot the confined band structure for a uniform 10 nm wide InAs contact region, obtained by rewriting Eq. (4.10) as in Eq. (3.23) and solving the resultant eigenvalue problem. Fig. 4.2 shows the subbands that are formed in both conduction and valence band. For a given energy, the subbands with an ingoing probability current can be injected into the device as discussed in Section 3.4. Thanks to the introduction of an effective mass in the orthogonal direction, the ordering of the valence band subbands is now correct, with the subband number increasing for more negative energy (see the arrow in Fig. 4.2).

Note that although the effective masses in Eqs. (4.9) and (4.10) result in a correct direction of curvature for the valence band in the orthogonal directions, they do not reintroduce an actual coupling between conduction and valence band in those directions. This means BTBT is only possible in x, the direction of transport. Our two-band model is therefore poorly suited to model configurations that have a significant BTBT component orthogonal to the gate. This is especially true for the lineTFET, in which the orthogonal BTBT is the main current component. But also pointTFET configurations in which FIQC underneath the gate dielectric plays an important role are not well modeled. There are, however, some interesting configurations which can already be studied with this basic two-band model, as will be discussed in Section 4.4.

Although the ordering of valence band subbands in Fig. 4.2 is correct, we see that the valence band still curves upwards for high k_x-values, from halfway the Brillouin zone towards the edge, creating states close to and inside the bandgap. These so-called spurious solutions are clearly unphysical, since the periodicity of the band structure in the different Brillouin zones is no longer satisfied. Before we can proceed with transport calculations, we first need to have a strategy to deal with these spurious solutions.

4.2.3 Spurious solutions

Spurious solutions are solutions of the truncated EF system, such as the one in Eq. (4.9), that are not solutions of the exact, infinite band system [144]. They are therefore a consequence of solving an approximate, limited band model.

The E-k diagram of a truncated EF system is no longer periodic over the different Brillouin zones, but rather shows an upwards curvature of the valence band at high k_x-values, as shown in Fig. 4.2. This behavior of the valence

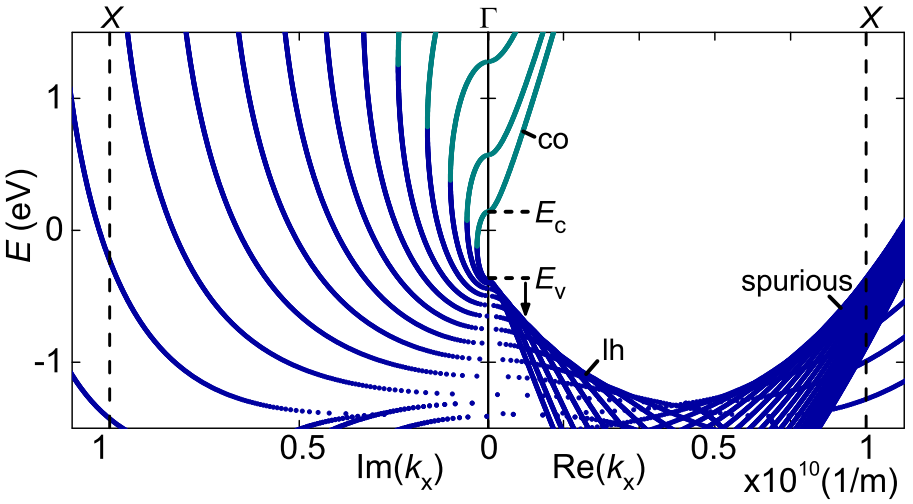

Figure 4.2: Real (right) and imaginary (left) two-band confined band structure of a 10 nm wide InAs slab in the [100] transport direction. k_y is zero. The arrow indicates rising valence band subband number. The dashed vertical lines indicate the edge of the first Brillouin zone. The solutions are colored based on the dominant basis state. Indicated are the conduction (co) and light hole (lh) bands.

band can be understood from the form of the eigenvalue problem Eq. (3.22) that was solved to obtain Fig. 4.2, and which is basically a rewritten version of Eq. (4.10). Eq. (3.22) is a quadratic equation in k_x, with a positive coefficient for the quadratic term. This means, as we have seen in Section 4.2.1, that any downwards curvature is the result of terms linear in k_x. In a model with a finite number of bands, and therefore a finite number of linear terms, the quadratic term will inevitably dominate the linear terms for sufficiently high k_x. This is true for all bands, including the valence bands. Once the quadratic term dominates, we essentially have the eigenvalue problem for a free electron, of which the dispersion is an upwards curving parabola.

Although we now know their origin, the question remains: what to do with these spurious solutions in transport calculations? One option is to simply neglect them by filtering all solutions above a chosen cut-off for k_x. However, Burt has shown that, although unphysical, the "out-of-zone" spurious solutions as he calls them, are required to ensure continuity of the envelope functions and their first derivatives at a heterojunction [144]. So although we do not inject the spurious modes, we do need to keep them in the boundary conditions as possible transmitted or reflected modes. In the discretized boundary conditions outlined

in Section 3.4.2, this means we filter spurious modes from $\mathbf{S}_{11}$ (ingoing source modes) and $\mathbf{D}_{12}$ (ingoing drain modes), but retain them in $\mathbf{S}_{12}$ (outgoing source modes) and $\mathbf{D}_{11}$ (outgoing drain modes). As a sanity check, we verify that the retained spurious modes do not carry any significant probability current, meaning that the corresponding coefficients in $\mathbf{r}$ and $\mathbf{t}$ are small.

Next to the out-of-zone solutions, Fig. 4.2 also shows imaginary subbands that cross the bandgap. As discussed in Section 2.5.1, these solutions are not spurious, but show the coupling between the different bands. These solutions are quickly decaying, which means they have a negligible transmission probability. To save computational time, we therefore handle them like the out-of-zone solutions: they are not injected, but kept as possible reflected or transmitted modes.

Now that we have a strategy for dealing with spurious solutions, we calibrate the parameters of the model.

4.3 Parameter calibration

Eq. (4.9) shows that our two-band model has a limited number of parameters for each material layer: p, E_c, E_v, m_c^* and m_v^*. p cannot be directly determined experimentally and is discussed in Section 4.3.1. E_c and E_v can be obtained from the electron affinity E_EA and bandgap E_G, which can be measured optically and are well known in literature for III-V materials [139]. Finally, m_c^* and m_v^*, which determine the orthogonal DOS, can be measured with cyclotron measurements and can be found in literature as well [139]. However, as discussed in Section 4.2.2, it is uncertain whether m_v^* should be the heavy or the light hole mass, although from theory we expect the heavy hole mass. This choice for m_v^* is discussed in Section 4.3.2.

4.3.1 Momentum matrix element

One way to calibrate the momentum matrix element p is through various fitting procedures as suggested in literature [139]. In these studies, p is varied in an eight-band model, such that certain target effective masses are obtained at the Γ-point. The resulting values for p are commonly listed in units of energy as:

$$E_\mathrm{P} = \frac{2p^2}{m_e} \tag{4.11}$$

Literature does not mention, however, whether these E_P-values are transferable to a two-band model. To determine E_P, we therefore carry out a calibration

Table 4.1: SC calibrated values for the parameters in the Kane model.

	Lower limit	Recommended value	Upper limit
A_{BTBT} [$\mathrm{cm^{-3}s^{-1}}$]	1.1x10^{20}	1.3x10^{20}	1.6x10^{20}
B_{BTBT} [$\mathrm{V\ cm^{-1}}$]	6.0x10^{6}	5.7x10^{6}	5.4x10^{6}

based on experimentally determined electrical characteristics of $In_{0.53}Ga_{0.47}As$ p-i-n diodes. We also make a theoretical assessment of the value of E_{P} in a two-band versus an eight-band model.

Experimental calibration

$In_{0.53}Ga_{0.47}As$ p-i-n diodes were fabricated at imec by Quentin Smets [41]. The diodes have a layered p^+-i-n^+ structure, which was grown by molecular beam epitaxy (MBE) and patterned by e-beam lithography. The contact regions are degenerately doped. Diodes with different lengths of the intrinsic region (T_{i} = 9 nm, 18 nm and 46 nm) were fabricated to allow for the BTBT calibration at different electric field strengths.

Before carrying out the calibration of E_{P}, the diode results were used by Smets *et al.* to calibrate SC BTBT parameters [41]. The SC simulator used was Synopsys Sentaurus Device (SDevice), which relies on the Kane model as discussed in Section 2.5.1, to model BTBT. From Eq. (2.11) it is clear that the Kane model has two important parameters: A_{BTBT} and B_{BTBT}. A_{BTBT} is the pre-exponential factor and hence determines the vertical shift of the exponential BTBT-based IV-curve, while B_{BTBT} is in the exponential, and determines the slope of the IV-curve. To calibrate these two parameters, doping profiles, determined using secondary ion mass spectrometry (SIMS) complemented with capacitance-voltage measurements (C-V), were first imported into SDevice. Next, the simulated IV-curves were fitted to the experimental results in the reverse bias region. In the fitting procedure, the conduction band valleys at the Γ-, X- and L-points were included for the calculation of the Fermi level position. The Jain-Roulston bandgap narrowing (BGN) model was also activated, since the contact regions are degenerately doped. The result of this calibration procedure was a set of three values for the BTBT parameters, based on the experimental uncertainties: a lower limit, an upper limit and a recommended value (see Table 4.1).

We now use the calibrated SC parameter set in Table 4.1 to calculate the QM parameter E_{P}. E_{P} is linked to B_{BTBT} with the relation [47]:

$$B_{\mathrm{BTBT}} = \frac{\pi E_{\mathrm{G}}^2 m_e}{4\hbar\sqrt{\frac{E_{\mathrm{P}} m_e}{2}}} \tag{4.12}$$

where E_{G} is 0.74 eV for $In_{0.53}Ga_{0.47}As$. We obtain a value for E_{P} of 13.5 eV for the lower limit, 15 eV for the recommended value and 16.5 eV for the upper limit. The corresponding values for p can be calculated with Eq. (4.11).

Theoretical assessment

The value for E_{P} can also be estimated from theory. In a two-band model without spin-orbit coupling or perturbative influence of other bands, E_{P} can be extracted from the following relation [139]:

$$\frac{m_e}{m_{\mathrm{c}}^*} = 1 + \frac{E_{\mathrm{P}}}{E_{\mathrm{G}}} \tag{4.13}$$

where an effective mass m_{c}^* of $0.043 m_e$ for $In_{0.53}Ga_{0.47}As$ results in a value for E_{P} of 16.5 eV, which is within the limits obtained from the calibration.

The value that we obtain is different from values cited in literature, where E_{P} has a recommended value of 25.3 eV [139]. This value, however, is calibrated for an eight-band model. Such a model consists of a conduction band, a heavy hole band and a light hole band, with the effect of higher bands included perturbatively through Luttinger parameters. It also includes the effects of spin. In such a model Eq. (4.13) is modified to

$$\frac{m_e}{m_{\mathrm{c}}^*} = 1 + 2F_{\mathrm{K}} + \frac{E_{\mathrm{P}}(E_{\mathrm{G}} + 2\Delta/3)}{E_{\mathrm{G}}(E_{\mathrm{G}} + \Delta)} \tag{4.14}$$

with F_{K} the Kane parameter, which includes the perturbative effect of higher bands, and Δ the spin-orbit splitting energy. It is clear that for a given m_{c}^*, the presence of the term with F_{K} requires a change in E_{P} compared to the two-band model. In the case of $In_{0.53}Ga_{0.47}As$, F_{K} is 1.77, while Δ is 0.33 eV, resulting in the eight-band E_{P} of 25.3 eV. Since the value of E_{P} determines the BTBT coupling strength, using the eight-band E_{P} would result in an artificially high BTBT rate. Hence, based on this analysis, we conclude that an eight-band model is not suitable for BTBT predictions.

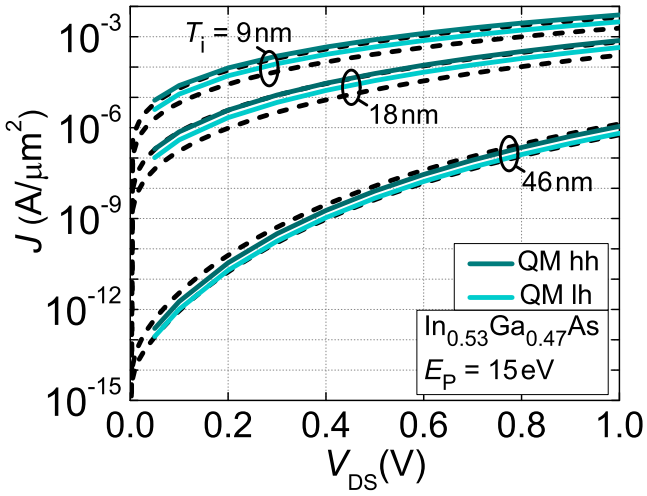

Figure 4.3: Simulated SC (dashed) and two-band QM (solid) current densities in reverse bias for p-i-n diodes with different lengths of the intrinsic region. The two-band simulations are carried out with a light hole (lh) and heavy hole (hh) effective mass in the orthogonal directions. The dashed curves correspond to SC simulations carried out with the upper and lower values for A_{BTBT} in Table 4.1, and with the BGN model deactivated. The n-type doping level is $2.2\mathrm{x}10^{19}\mathrm{cm}^{-3}$ and the p-type doping level is $1.7\mathrm{x}10^{19}\mathrm{cm}^{-3}$.

4.3.2 Orthogonal effective mass

With the calibrated E_{P}, we can now assess which valence band effective mass to choose for the orthogonal directions in the two-band EF system of Eq. (4.9). Fig. 4.3 shows QM simulations with the calibrated E_{P} and the two options for m^*_{v}. We see that both the heavy and the light hole mass fall within an acceptable range of the calibrated SC simulations. We therefore retain our choice for the heavy hole mass when using the two-band model in the simulations of Section 4.4, as it is expected from theory. We verify this again with a fifteen-band model in Section 5.5.3.

4.4 Application to heterostructure diodes and TFETs

In this section, we use our two-band QM implementation and the calibrated parameters of the previous sections to simulate heterostructure diodes (Section 4.4.2) and TFETs (Section 4.4.3). The aim is to investigate the effect of SIQC on the transmission spectrum and the current, an assessment that can not be performed with a semiclassical solver.

4.4.1 Simulation details

We investigate lattice-matched heterostructure diodes with body thicknesses (T_{body}) ranging from wide (60 nm) to thin (5 nm). Two heterojunction alignments are considered: a straddled alignment ($In_{0.53}Ga_{0.47}As/InP$) and a staggered alignment ($GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$). Fig. 4.4 shows the band edge alignments at the tunnel junction in more detail. It is assumed that the alignment is only determined by the electron affinities of the materials. Further details of the simulated diode configurations can be found in Fig. 4.5(a) and Table 4.2. The two-band QM parameters can be found in Appendices C.2.1 ($In_{0.53}Ga_{0.47}As$), C.2.2 ($GaAs_{0.5}Sb_{0.5}$) and C.2.3 (InP).

In a next step, we investigate the impact of T_{body} for heterostructure pocketed pointTFETs with a staggered $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ band alignment at the tunnel junction (see Fig. 4.5(b)). We only study the staggered alignment, as it is the most promising for TFET applications because of the small effective bandgap at the tunnel junction (see Section 2.7.3). We choose a configuration

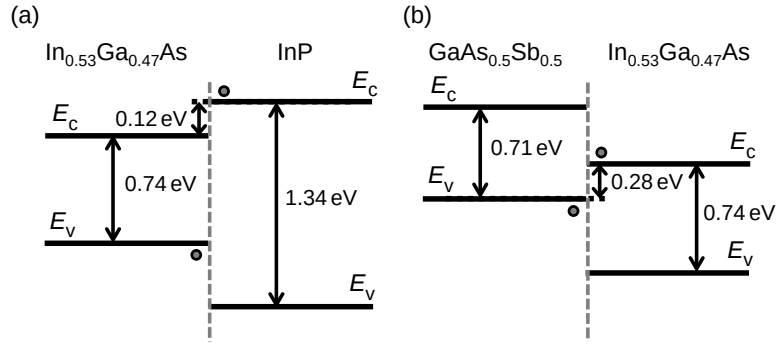

Figure 4.4: The band edge alignments of the studied heterostructures with (a) a straddled and (b) a staggered alignment. The circles indicate the bands between which the tunneling transition takes place.

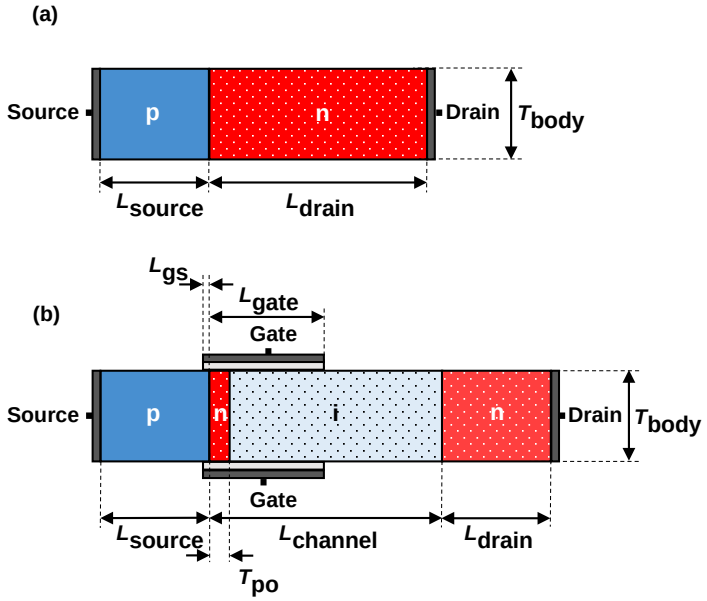

Figure 4.5: The simulated configurations of (a) heterostructure diodes and (b) heterostructure pocketed pointTFETs.

Table 4.2: Details of simulated configurations.

	Diode	TFET
L_{source} [nm]	10	10
L_{channel} [nm]	/	20
L_{drain} [nm]	20	10
$L_{\text{gate}}/L_{\text{gs}}$ [nm]	/	22/1
T_{body} [nm]	varying	varying
T_{po} [nm]	varying	varying
Doping source [cm^{-3}]	5x10^{19}	5x10^{19}
Doping drain [cm^{-3}]	5x10^{19}	1x10^{19}
Doping pocket [cm^{-3}]	/	5x10^{19}
Equivalent oxide thickness (EOT) [nm]	/	0.6
Gate work function (GWF) [eV]	/	4.05

with a counterdoped pocket at the source region, since this forces tunnel paths more parallel to the gate, as discussed in Section 2.8.1. The absence of perpendicular tunnel paths in the two-band model therefore does not compromise the predictions for this type of device. In addition, the increased electric field at the tunnel junction is expected to improve I_{ON} and reduce the undesired SS stretch-out of FIQC. T_{po} is chosen slightly larger (2 nm) than the depletion width of the junction (1.9 nm), which is the optimal T_{po} as discussed in Section 2.8.1. Note that for the thinnest configuration with a T_{body} of 5 nm, a gate-source underlap of 1 nm was introduced to prevent depletion across the body in the source region (see Section 2.9.1). Other details of the configurations are provided in Fig. 4.5(b) and Table 4.2.

Our implementation contains both a spatial and an energetic mesh. The spatial directions are uniformly discretized with steps of 0.5 nm. The energy range is equal to the tunneling window, bounded by the valence band edge in the source and the conduction band edge in the drain. This range is uniformly discretized into 40 points. We have verified that the results are insensitive to a further refining of both the spatial and the energetic mesh.

The external electrostatic potential is obtained from SC simulations with SDevice. This simulator solves Poisson's equation self-consistently with the electron and hole continuity equations, which contain terms for carrier drift and diffusion, but not BTBT. The SC simulator can incorporate an arbitrary number of bands and conduction band satellite valleys in an effective mass approximation. We therefore have to make a choice: either include only the two bands that are also present in the QM transport calculation, or include all bands for which effective mass parameters are available. We opt for the latter, since this leads to a more accurate position of the Fermi level. To further improve the energetic position of the Fermi level in the conduction band, a non-parabolicity correction is activated. Our approach is therefore to obtain an electrostatic potential as accurate as semiclassically possible, and then use this potential in the QM simulations. An example SDevice input script has been included in Appendix D, along with the material parameter files, containing the effective masses and non-parabolicity correction factors for the considered materials in Appendices D.3.2 ($In_{0.53}Ga_{0.47}As$), D.3.3 ($GaAs_{0.5}Sb_{0.5}$) and D.3.4 (InP).

It is important to note that Poisson's equation is not solved self-consistently with the QM calculation of the carrier density. Quantum effects on the distribution of the carrier density, e.g. due to FIQC or SIQC, are therefore not reflected in the electrostatic potential. In the channel of the simulated TFETs, the impact of the carrier redistribution on the electrostatic potential is small if the carrier density is not too high. In the contacts of both the diodes and the TFETs, on the other hand, the Fermi level position is that of the non-confined case. The simulated active carrier density is therefore lower in the simulated device

because of SIQC. For the diodes, the absence of SIQC also means the potentials are constant in the confined direction, regardless of T_{body}. Hence, care should be taken in the interpretation of the results for strongly confined configurations. We return to this issue in Chapter 7, where we develop a fully self-consistent simulation procedure.

4.4.2 Quantum confinement effects in heterostructure diodes

Figs. 4.6 and 4.7 illustrate the effects of SIQC for the straddled InGaAs/InP diodes. Fig. 4.6 shows that due to confinement, a discrete set of subband modes is formed in the contacts, as was displayed schematically in Fig. 3.2. For decreasing T_{body}, the separation between the subband energy levels increases, such that the forbidden bandgap seen by the traveling modes enlarges. The larger bandgap decreases the transmission probability of the modes, as shown in Fig. 4.7. For the 5 nm diode, the confinement is so strong that only one mode has a significant transmission probability. The bandgap for the other modes has increased to a point where the transmission probability becomes negligible.

It is interesting to note that the transmission spectrum of the dominant mode of the 5 nm diode coincides with the second mode of the 10 nm configuration. This is a result of the effective mass approximation that was introduced in Section 4.2.1 for the orthogonal directions. Since it is the orthogonal effective mass that determines the confined subband energy levels, the energy of the nth subband level of conduction or valence band can be estimated as [145]:

$$E_{\text{n}} = \frac{n^2 \hbar^2 \pi^2}{2 m^*_{\text{c,v}} T^2_{\text{body}}} \tag{4.15}$$

with $m^*_{\text{c,v}}$ either the effective mass of the conduction or valence band. From Eq. (4.15), it can be verified that E_1 for a T_{body} of 5 nm is expected to coincide with E_2 for a T_{body} of 10 nm. The bandgap that is seen by the first subband mode of the 5 nm diode is therefore the same as that seen by the second subband mode of the 10 nm configuration, resulting in identical transmission spectra.

In the remainder of this chapter, we will normalize the transmission probabilities to T_{body}, which results in a transmission probability per unit of T_{body}. This isolates the effect of confinement on the subbands themselves, rather than on the amount of subbands available. Figs. 4.8 and 4.9 then show the effect of SIQC on the normalized transmission probability as T_{body} of the diodes is reduced. This decrease of transmission probability is reflected in the current densities of Figs. 4.10 and 4.11. The effect of SIQC on both the transmission spectrum and the current density becomes significant for T_{body} below 30 nm for these particular material systems.

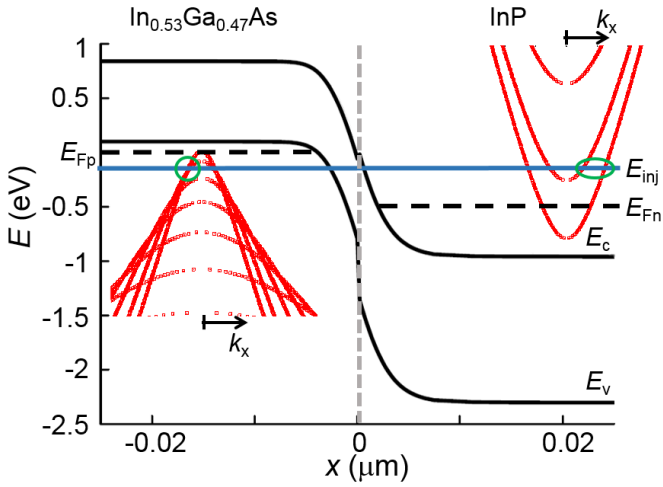

Figure 4.6: Energy bands (black) and quantized contact subband structure for a 5 nm InGaAs/InP heterostructure diode at a V_{DS} of 0.5 V. The circles indicate right propagating subband modes available for transport at a given injection energy level (E_{inj}). The grey dashed line indicates the heterojunction. The black dashed lines indicate the quasi-Fermi levels of electrons (E_{Fn}) and holes (E_{Fp}). Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

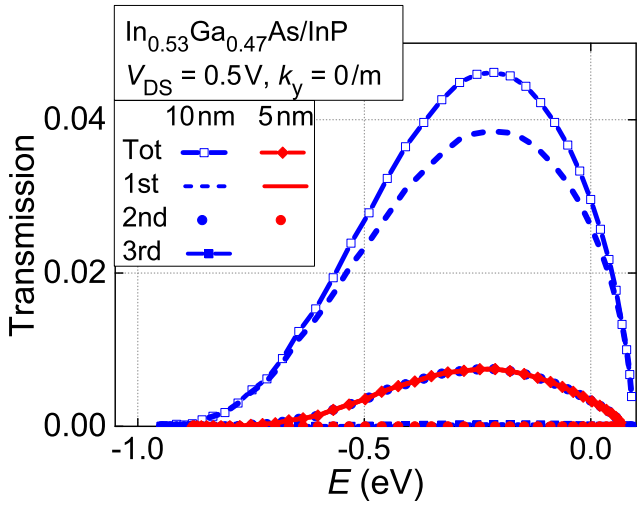

Figure 4.7: Simulated transmission spectrum for the individual propagating modes in a straddled heterostructure diode with a T_{body} of 10 nm and 5 nm. Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

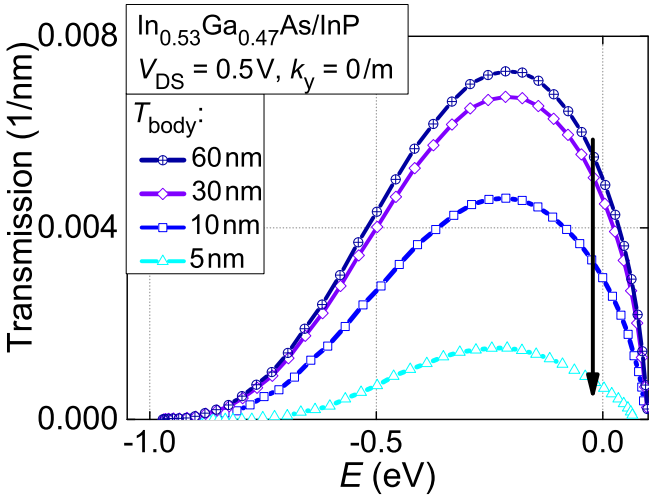

Figure 4.8: Simulated normalized transmission spectrum in a straddled heterostructure diode for varying degrees of confinement. Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

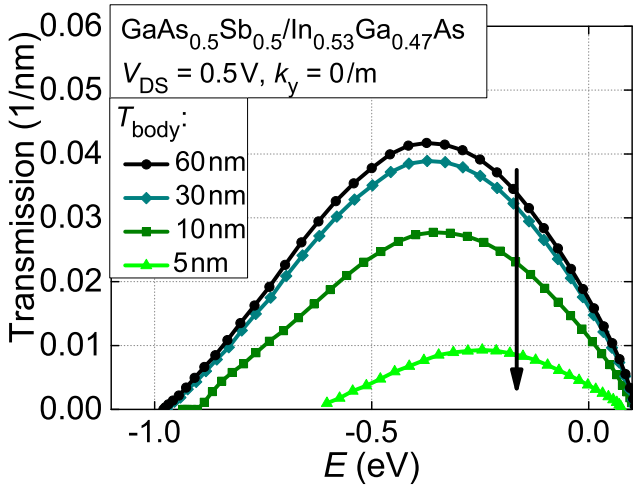

Figure 4.9: Simulated normalized transmission spectrum in a staggered heterostructure diode for varying degrees of confinement. Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

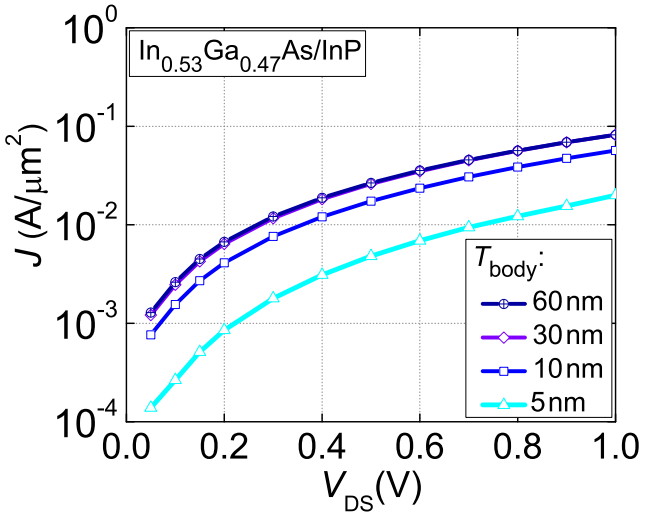

Figure 4.10: Simulated current density in a straddled heterostructure diode for varying degrees of confinement. Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

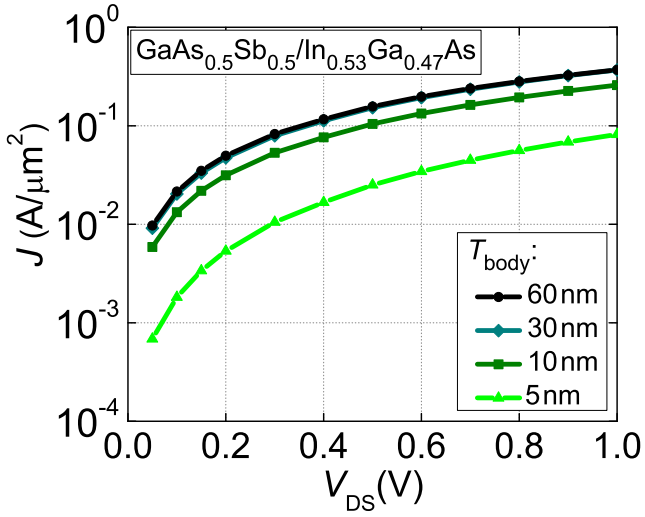

Figure 4.11: Simulated current density in a staggered heterostructure diode for varying degrees of confinement. Other configuration details are listed in Fig. 4.5(a) and Table 4.2.

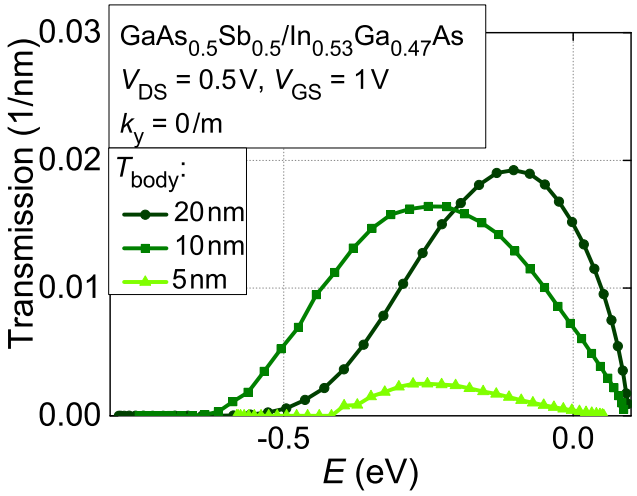

Figure 4.12: Simulated normalized transmission spectrum in a staggered heterostructure pocketed pointTFET for varying degrees of confinement. Other configuration details are listed in Fig. 4.5(b) and Table 4.2.

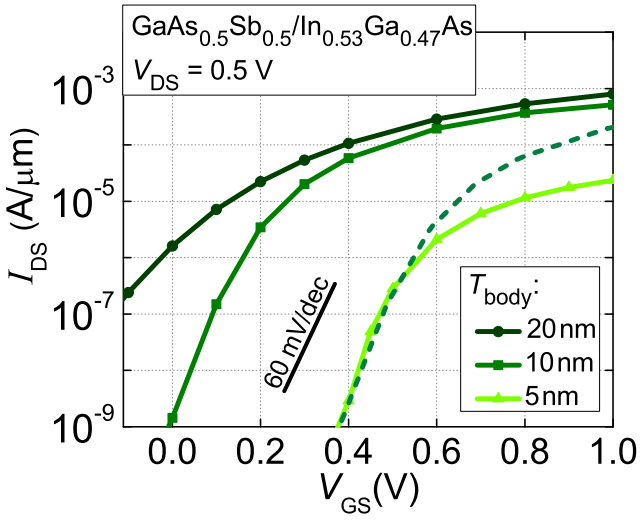

Figure 4.13: Simulated transfer characteristics in a staggered heterostructure pocketed pointTFET for varying degrees of confinement. The dashed line is the 10 nm curve shifted to the same I_{OFF} as the 5 nm curve. The simulated TFETs are double gate devices and the displayed currents are always for both gates. Other configuration details are listed in Fig. 4.5(b) and Table 4.2.

4.4.3 Quantum confinement effects in heterostructure TFETs

Figs. 4.12 and 4.13 show that the detrimental effects of SIQC, also present in diodes, are counteracted by an improved control of the gate over the tunnel junction for the thinner configurations. The improvement in performance as a result of better gate control was discussed in Section 2.9.1, and we encounter it here as T_{body} is scaled below 20 nm. For the 20 nm configuration, the gate control is weak, resulting in a degraded SS. As evident in Fig. 4.13, the tighter gate control improves the SS for the 10 nm and 5 nm configurations, although SIQC decreases I_{ON} for the 5 nm case. The same conclusion can be made based on the transmission spectra in Fig. 4.12. The transmission probability drops sharply for the strongly confined 5 nm case, while the improved gate control prevents the SIQC-induced degradation in the 10 nm case. Note, however, that the 5 nm case is not self-consistent. Due to the strong SIQC, changes in Fermi level position and carrier concentrations are expected to be non-negligible. Hence, the 5 nm case data are not too reliable.

4.5 Conclusion

In this chapter, we implemented our EF formalism of the previous chapter with a simple two-band model. The basis set of this model consists of one conduction band state and one valence band state. We showed that such a small basis set comes with important limitations, most notably the lack of orthogonal coupling between the bands. This leads to an upwards curvature of the valence band and a lack of BTBT in the orthogonal directions. Although we were able to reintroduce the correct curvature with an effective mass model, the lack of orthogonal BTBT severely restricts the types of configurations that can be accurately studied. After calibrating the model with $In_{0.53}Ga_{0.47}As$ diodes, we simulated heterostructure diodes and TFETs with straddled and staggered band alignments for varying T_{body}. For decreasing T_{body}, we found a counteracting effect between SIQC-induced current degradation and improved gate control. This suggests an optimum for T_{body} exists.

Because of the lack of orthogonal coupling, we were only able to simulate pocketed pointTFETs: a comparison with a lineTFET configuration, which depends on orthogonal tunneling, was not possible. In the next chapters, we surpass this limitation by extending to band models with a larger basis set. We will also verify that the two-band model indeed provides a good first approximation for non-confined pocketed pointTFETs.

The results in this chapter are published in the *Journal of Applied Physics* [142].

Chapter 5

Fifteen-band model

The two-band model of the previous chapter was only suited for the simulation of specially-designed pointTFETs, because of the lack of orthogonal BTBT. This prevented us from achieving the second goal of this thesis, which is to assess the viability of TFET in a direct bandgap III-V material system, including configurations like the lineTFET. In this chapter, we therefore extend the band structure model to fifteen bands. A fifteen-band model provides BTBT in all directions, which is essential to predict the lineTFET performance. It also provides a general improvement in the accuracy of the TFET predictions as it captures the full first Brillouin zone of the band structure, except for the features caused by spin-orbit coupling.

The upgrade to a fifteen-band model does come with an increased computational burden and more difficult to handle spurious solutions. We therefore implement a spectral method to alleviate these issues, before we can start the comparison and optimization of the III-V TFET configurations. In the next chapter, we go one step further and include the spin-orbit interaction and the influence of strain on the band structure, which allows to further explore the TFET design space.

The chapter is structured as follows. Similar to our discussion of the two-band model, we first introduce the fifteen-band basis set with its various parameters (Section 5.1). Next, we show how the spurious solutions of the model and the computational complexity prevent the use of the same finite difference implementation of the two-band case (Section 5.2). We circumvent these issues with a spectral approach to the solution of the EF equations (Section 5.3). We then develop a calibration procedure to obtain fifteen-band parameters for any material, while retaining commutativity of the momentum operator (Section 5.4).

With the spectral implementation and the calibrated parameters, we can simulate a wide range of TFET configurations. First, we compare pocketed homojunction pointTFET and lineTFET n-channel configurations (Section 5.5). In this section, we also compare the results with two-band simulations to show the necessity of the fifteen-band model (Section 5.5.3). At the same time, we verify the assumptions made when calibrating BTBT in the previous chapter regarding the orthogonal effective mass in the valence band. Next, the comparison between lineTFET and pointTFET is repeated for p-channel configurations, which typically perform worse than their n-channel counterparts. To enable complementary circuit implementations, we therefore introduce an improved source design that greatly enhances the pTFET performance and brings it at a similar level as the nTFET (Section 5.6). At the end of the chapter, we present results of an optimization of heterojunction TFETs, carried out in the imec TFET team with our fifteen-band implementation, with the aim of finding out whether TFETs can compete with MOSFETs (Section 5.7). Finally, we compare fifteen-band QM simulation results to SC simulations and experimental homostructure p-i-n results (Section 5.8). The findings of this chapter are summarized in the conclusion (Section 5.9).

5.1 Fifteen-band basis set

In the fifteen-band model, the wave function is expanded on a set of fifteen basis functions:

$$\psi(\mathbf{r}) = \sum_{n}^{15} F_n^k(\mathbf{r}) U_n^k(\mathbf{r}) \tag{5.1}$$

As in the two-band model, the basis set is listed as a set of basis states, shown in Fig. 5.1 with the associated energy levels and interband momentum matrix elements, which couple the different states [58]. The basis functions are then the projections of these states onto the position basis as in Eq. (4.2). For the valence band, the set includes one valence band state with s-type symmetry ($|S_V\rangle$) and three degenerate states with p-type symmetry ($|X\rangle$, $|Y\rangle$ and $|Z\rangle$), which are named after their respective directionalities. In a bulk band structure, these p-type valence band states correspond in each direction to a light hole band and two degenerate heavy hole bands. For the conduction band, the set contains three s-type states ($|S\rangle$, $|S_u\rangle$ and $|S_q\rangle$), the first of which corresponds to the lowest conduction band. The energetic distance between the energy of $|S\rangle$ and $|X\rangle$, $|Y\rangle$, $|Z\rangle$ is therefore the bandgap E_G. Additionally, the conduction band contains a set of three degenerate p-type states ($|X_c\rangle$, $|Y_c\rangle$, $|Z_c\rangle$) and five d-type states, which can be subdivided into a set of three ($|X_d\rangle$, $|Y_d\rangle$ and $|Z_d\rangle$) and two degenerate states ($|D_x\rangle$ and $|D_z\rangle$). These fifteen states are coupled

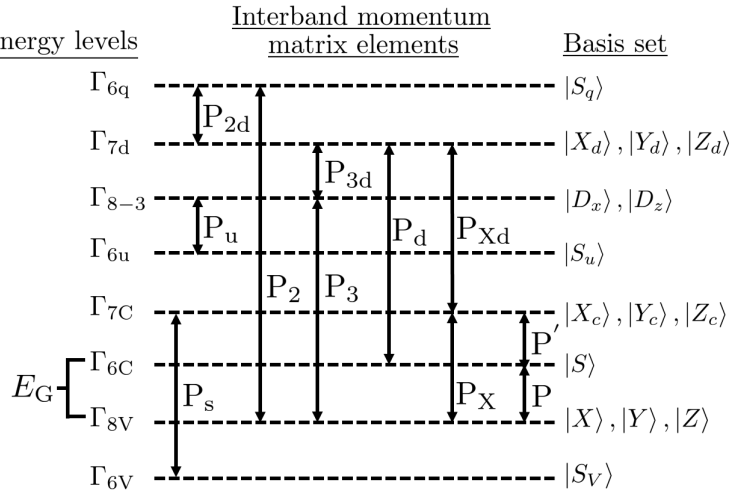

Figure 5.1: Energy levels and interband momentum matrix elements of the chosen fifteen-band basis set. This figure is based on Fig. 1 in [58], with corrected P_S.

with eleven interband momentum matrix elements and are associated with eight energy levels. Compare this to the two-band model, which includes only one conduction band state $|S\rangle$, coupled with a sole valence band state $|X\rangle$ through a single matrix element (see Chapter 4).

This choice of fifteen-band basis set provides a so-called "full-zone" description of the band structure. Full-zone means that it is able to correctly capture the band structure (without spin-orbit coupling) in the full first Brillouin zone. An example bulk band structure is shown in Fig. 5.2, illustrating that the X- and L-valleys are included in the model. Notice that outside of the first Brillouin zone, the bands all curve upwards similar to the two-band case, since the basis set is still finite (see also Section 4.2.3).

The fifteen-band basis set is also self-contained, which means all of the coupling occurs between states that are a member of the set. This is in contrast to e.g. an eight-band model, in which Luttinger parameters perturbatively incorporate the effect of states that are outside of the chosen set [139]. As we have seen in Section 4.3, this perturbative approach artificially increases the interband coupling parameter, which unphysically changes the BTBT strength.

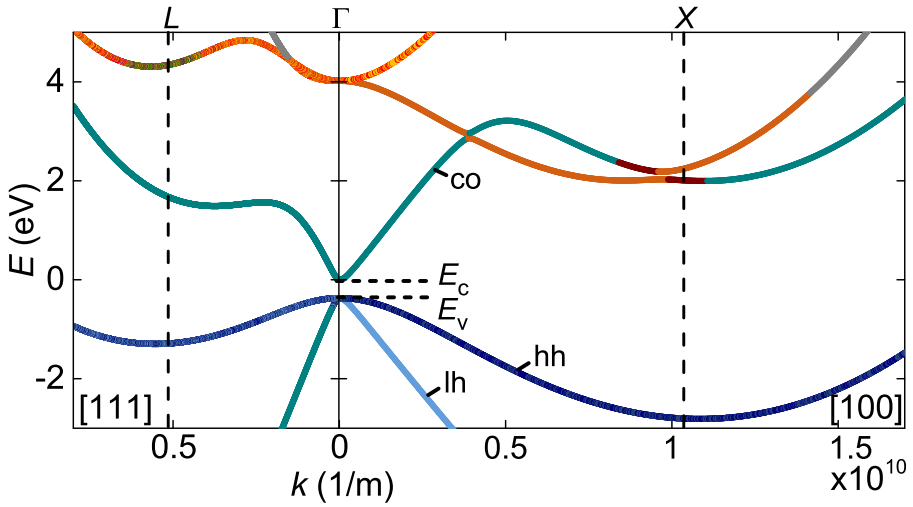

Figure 5.2: Fifteen-band bulk band structure of InAs in the [100] transport direction (right) and the [111]-direction (left). The dashed lines indicate the edge of the first Brillouin zone. The solutions are colored based on the dominant basis state. Indicated are bands which can be associated to the conduction (co), light hole (lh) and heavy hole (hh) bands. The fifteen-band parameters are taken from Radhia *et al.* [58] and are listed in Appendix C.2.

5.2 Model-specific issues

Unfortunately, a straightforward introduction of the fifteen-band expansion of Eq. (5.1) into Eq. (3.15) introduces specific issues, such as a significant increase in computational burden (Section 5.2.1) and the appearance of spurious solutions which pollute the bandgap (Section 5.2.2).

5.2.1 Computational complexity

If we take the same approach as in the two-band case and discretize the EF system of Eq. (3.15) in a fifteen-band model with a finite difference method, the computational burden would become prohibitively high. As discussed in Section 3.4, the size of the discretized system would be $(N_x \cdot N_z \cdot N \times N_x \cdot N_z \cdot N)$. Assuming the same mesh as for the two-band diode and TFET simulations in Section 4.4, $\mathbf{K}_{\text{tot}}$ in Eq. (3.49) would have dimensions 540 000 × 540 000. Considering that the LU-factorization used to solve this system has a

time complexity of $\mathcal{O}(n^3)$ for an $n \times n$ matrix, it is clear that the computational cost quickly becomes very high.

5.2.2 Spurious solutions

Even if the computational resources would be available to solve the fifteen-band finite difference system, any result for a confined structure is plagued by spurious solutions. As we can see in the confined band structure of Fig. 5.3, the spurious solutions are not just out-of-zone solutions at high values of k_x as in the two-band model or the bulk fifteen-band model. Rather, additional real and complex spurious solutions now also pollute the band structure around the zone center. The result is the absence of a bandgap.

The out-of-zone solutions can be easily recognized based on their high k_x-value and can therefore be handled as discussed in Section 4.2.3 for the two-band case. The zone-center spurious solutions, on the other hand, are much harder to distinguish from the physical solutions. This makes it difficult to assess which of the corresponding modes should be injected in a transport calculation and which should not.

The zone-center spurious solutions originate from out-of-zone solutions that are mapped back onto the zone-center when the structure is confined in one of the orthogonal directions. From the bulk band structure in Fig. 5.2, we know that the out-of-zone spurious solutions, caused by the upwards curvature of the valence band, are present in all directions, including the k_z-direction. As discussed in Section 4.2.3, when the structure is confined in the z-direction, the effect on the band structure can be pictured as taking slices along the k_z-axis, followed by projecting these slices onto the $k_z = 0$ plane (see Fig. 4.1). For high values of k_z, the valence bands of the fifteen-band model curve upwards, and as a result, the slices of these out-of-zone solutions are projected into the zone-center, and eventually into the bandgap. Note that we do not encounter this issue for the two-band model, since there we implement an effective mass model in the confined z-direction, which is a parabolic approximation with no upwards curvature of the valence band. In Section 5.3, we use the insight into the origin of the zone-center solutions to develop a strategy to remove them from the band structure.

Next to the out-of-zone and zone-center solutions, Fig. 5.3 also shows several complex solutions. These have been described in literature as an inherent property of higher band models [146, 147]. In contrast to the purely real zone-center solutions, complex spurious solutions can be easily distinguished based on their large imaginary part. We apply the same strategy as for the out-of-zone solutions, described in Section 4.2.3: we do not inject the modes corresponding

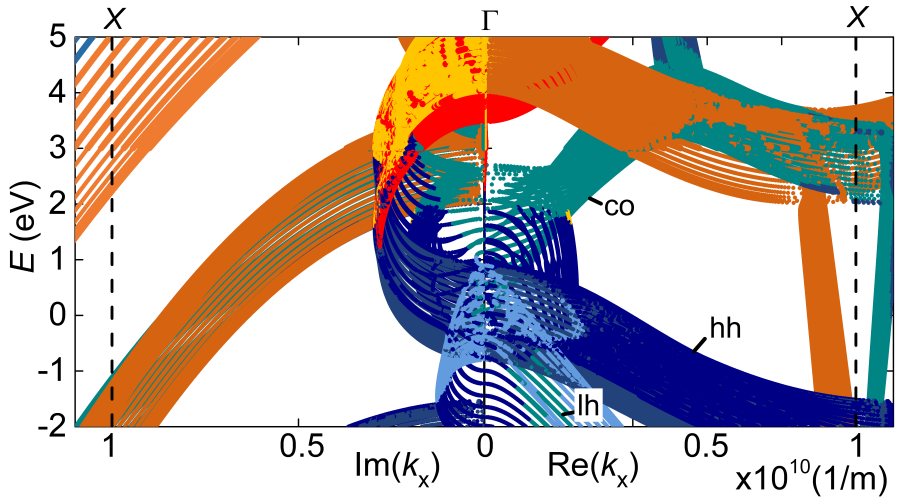

Figure 5.3: Fifteen-band confined real (right) and imaginary (left) band structure of a 10 nm InAs slab in the [100] transport direction, calculated using finite differences. The dashed vertical lines indicate the edges of the first Brillouin zone. k_y is zero. The solutions are colored based on the dominant basis state component in the EF vector. Indicated are bands which can be associated to the conduction (co), light hole (lh) and heavy hole (hh) bands. The fifteen-band parameters are taken from Radhia *et al.* [58] and are listed in Appendix C.2.

to complex values of k_x, but do keep them as possible reflected or transmitted modes.

5.3 Spectral method for confined direction

To circumvent the issues with computational complexity and spurious solutions outlined in the previous section, we replace the finite difference discretization of the EF system in the confined direction with a spectral approach. In this section, we first introduce spectral methods in general, with their advantages and disadvantages (Section 5.3.1). Next, we discuss our choice of spectral functions (Section 5.3.2), and the resulting spectral EF system (Section 5.3.3). We then determine the associated discrete transformations (Section 5.3.4) and illustrate their use with an example (Section 5.3.5). Finally, we discuss a band structure calculated with the spectral implementation (Section 5.3.6).

5.3.1 Introduction

A spectral method is a technique to solve differential equations by expanding the solution functions, in our case the EFs, on a known, complete set of orthogonal spectral functions [148]. The problem is thereby changed from finding a value for the solution functions at each mesh point to finding the coefficients of these spectral functions. A choice has to be made for the spectral functions, to best suit the application at hand.

Numerically, a spectral method is implemented using the discrete transform associated with the chosen spectral function set. The exact form and properties of the transform depend on the choice of spectral functions. With this transform, the system to solve is transformed from the original domain (e.g. the position domain) to the spectral domain (e.g. the frequency domain). For a properly chosen spectral function set, the size of the system in the spectral domain is typically much lower than in the original domain. The transformed system is then solved and if desired, the reverse transformation brings the solutions back into the original domain.

Spectral methods have both advantages and disadvantages compared to finite difference/finite element methods (FD/FE). Here we list the ones most relevant for our application:

Advantages:

- **Exponential accuracy:**
  Spectral functions are defined on the full, global domain of the differential equation, in our case the full confined device width. FD/FE methods, in contrast, employ a local approximation. The global nature of the spectral decomposition results in so-called "exponential accuracy", compared to polynomial accuracy for FD/FE methods. This means that the addition of one function to a set of spectral functions reduces the relative approximation error exponentially, while for FD/FE methods the error scales with a polynomial of the step size. Only a small number of spectral functions is therefore typically required to achieve the same accuracy as FD/FE for the solution function, and hence the resulting set of equations can be small. The computational gain from the exponential accuracy usually trumps the loss in matrix sparsity that comes with the global nature of the spectral functions.

- **Efficient implementations:**
  Efficient numerical implementations exist for several spectral transformations, e.g. the fast Fourier transform (FFT), which scale as $\mathcal{O}(n\log(n))$.

The additional step of transforming the system to solve to the spectral domain therefore does not incur a large computational penalty.

- **Frequency selection:**
  If the chosen spectral domain corresponds to a physical quantity such as frequency, a limitation in the spectral domain corresponds to a selection of certain frequencies in the solution functions. This allows, for instance, to retain only the low or the high frequency components of the solution functions.

Disadvantages:

- **Inflexible device domain:**
  Because of the global nature of the spectral functions, the possible shapes of the device domain that can be modeled are determined by the choice of spectral function set. For a different type of shape (e.g. a circular domain instead of a rectangular), a change in the type of spectral functions is required. With FD/FE methods, on the other hand, a variety of shapes can be modeled without large changes to the implementation.

- **Gibbs oscillations:**
  Spectral methods are not well suited to capture abrupt local variations, such as discontinuities, in the approximated functions. In such cases, a very high number of spectral functions is needed to capture the abrupt change, resulting in the appearance of Gibbs oscillations in the solution. These are local overshoots at the discontinuities, of which the height does not decrease with increasing number of spectral functions [149].

For our application to the fifteen-band implementation, the advantages of spectral methods are very attractive. The exponential accuracy and efficient implementations give the prospect of a strong reduction in computational burden, providing the EFs and the potential do not vary abruptly. As discussed in Section 2.5.2, this is an inherent assumption of the EF formalism. We therefore expect only a small number of properly chosen spectral functions to be required. Furthermore, the possibility of frequency selection provides a handle to exclude the zone-center spurious solutions, which are highly oscillating because of their out-of-zone origin.

5.3.2 Choice of spectral functions

The previous section showed that spectral methods have several advantages which are interesting for application to the fifteen-band model. However, to benefit maximally from these advantages while maintaining the desired boundary conditions, the choice of spectral functions is crucial. Several options exist, each suited for specific applications: Fourier series, sine/cosine series, Chebyshev series, Laguerre series... (for an extensive list, see Appendix A of [148]).

In our implementation, we choose to expand the EFs in the confined z-direction onto a combined set of sines and cosines. The motivation to choose sines comes from the two-band contact eigenvalue problem Eq. (4.10): in the case of a constant potential energy in the confined z-direction, the eigenfunctions are a set of sines (it is then essentially a "particle-in-a-box" problem). In general, the potential is not constant in the confined direction, especially in the active device region, but if the variation in the z-direction is not too strong, we expect only a small set of sines to be required to reproduce the actual EFs in the z-direction. In our fifteen-band implementation, however, we need both sines and cosines, because in contrast to the two-band model, the EFs are coupled in the EF system through a derivative term in z (see Eq. (3.15)):

$$-\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x,z)\frac{\partial F_m^k(x,z)}{\partial z} \tag{5.2}$$

This means that if a given EF, say F_n, is decomposed into a set of sines, all of the EFs F_m to which it couples through the linear derivative term in z, have to be decomposed into a set of cosines. This effectively divides the set of EFs in odd (sines) and even (cosines) functions. We can then write the spectral decomposition as follows:

$$F^{(o)}(x,z) = \sum_{\mu=-\infty}^{+\infty} \tilde{F}^{(o)}(x,k_{z\mu})\sin(k_{z\mu}z) \tag{5.3}$$

$$F^{(e)}(x,z) = \sum_{\mu=-\infty}^{+\infty} \tilde{F}^{(e)}(x,k_{z\mu})\cos(k_{z\mu}z) \tag{5.4}$$

where the tilde indicates the spectral components, the superscripts o and e denote odd and even EFs respectively and $k_{z\mu} = \mu\frac{\pi}{L_z}$ with L_z the width in the z-direction. The index μ runs from $-\infty$ to ∞, making the expansions overcomplete. This is required for the derivation of the spectral EF system in Section 5.3.3 and Appendix B. In a numerical implementation, μ is limited based on a trade-off between desired accuracy and computational burden.

Important to note is that the type of spectral function determines the boundary conditions in the z-direction. This will be discussed in more detail for the discretized system in Section 5.3.4, but for now it is sufficient to know that the sine functions result in Dirichlet (zero value) boundary conditions, while the cosine set allows for any boundary value. Note that although therefore the even EFs are not necessarily zero at the boundaries, the Dirichlet boundary condition is still satisfied on the full wave function, because the basis functions have the opposite symmetry of the EFs [150]. This means that for each even EF, the corresponding basis function is odd and becomes zero on the boundary.

Besides the EFs in the active area, also the potential energy function and the confined envelopes in the contacts need to be expanded on the spectral set. The potential energy can vary along the boundary of the device. We therefore choose a set of cosines, since we do not want to force the potential to be zero at every point of the boundary in the confined direction:

$$V_{\mathrm{e}}(x,z) = \sum_{\mu=-\infty}^{+\infty} \tilde{V}_e(x,k_{z\mu})\cos(k_{z\mu}z) \tag{5.5}$$

For the confined envelopes in the contacts, the decompositions are analogous to those of the EFs in the active area in Eqs. (5.3) and (5.4):

$$\chi_\alpha^{(o)}(z) = \sum_{\mu=-\infty}^{+\infty} \tilde{\chi}_\alpha^{(o)}(k_{z\mu})\sin(k_{z\mu}z) \tag{5.6}$$

$$\chi_\alpha^{(e)}(z) = \sum_{\mu=-\infty}^{+\infty} \tilde{\chi}_\alpha^{(e)}(k_{z\mu})\cos(k_{z\mu}z) \tag{5.7}$$

where we again need both sines and cosines because of a coupling term similar to Eq. (5.2).

5.3.3 Spectral EF equations

Now that we have chosen a set of spectral functions, namely a set of sines and cosines, the next step is to transform the system of EF equations to the domain of the chosen spectral set. We do this both for the active area EF system of Eq. (3.15) and the contact eigenvalue problem of Eq. (3.19). Inserting

Eqs. (5.3)-(5.5) into Eq. (3.15), we obtain the spectral EF system:

$$\begin{aligned}
&\frac{-\hbar^2}{2m_e}\frac{\partial^2 \tilde{F}_n^k(x,k_{z\mu})}{\partial x^2} + k_y^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu}) + k_{z\mu}^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu}) \\
&- \frac{i\hbar}{m_e}\sum_m p_{nm}^{k,x}(x)\frac{\partial \tilde{F}_m^k(x,k_{z\mu})}{\partial x} + k_y\frac{\hbar}{m_e}\sum_m p_{nm}^{k,y}(x)\tilde{F}_m^k(x,k_{z\mu}) \\
&\pm k_{z\mu}\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x)\tilde{F}_m^k(x,k_{z\mu}) + \sum_m H_{nm}^k(x)\tilde{F}_m^k(x,k_{z\mu}) \\
&+ \sum_{\mu'=0}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu}) \mp \tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu'}) = E\tilde{F}_n^k(x,k_{z\mu}),
\end{aligned} \tag{5.8}$$

where the top (bottom) sign corresponds to the odd (even) EFs. We have assumed any heterojunction to be in the x-direction. The full derivation of this system is given in Appendix B, with the momentum matrices listed in Appendix A.4. Note the potential energy term, which is now a combination of a convolution and an autocorrelation.

The numerical size of Eq. (5.8) is determined by the number of spectral components, which is typically much lower than the amount of discretization points in a finite difference approach. For N_{k_z} spectral components, μ runs from 0 to $N_{k_z}-1$ instead of from $-\infty$ to ∞ as in Eqs. (5.3) and (5.4), which results in a discretized system of dimensions $N_x \cdot N_{k_z} \cdot N \times N_x \cdot N_{k_z} \cdot N$. Thanks to the exponential accuracy of the spectral method, N_{k_z} can be much smaller than N_z, especially when the EFs do not vary strongly in the confined z-direction. The value of N_{k_z} will be further discussed in Section 5.3.4.

Through a similar derivation, in which Eqs. (5.6) and (5.7) are inserted into the QTBM eigenvalue problem of Eq. (3.19), we obtain the following spectral

eigenvalue problem in the contacts:

$$
\begin{aligned}
&k_{x,\alpha}^2 \frac{\hbar^2}{2m_e} \tilde{\chi}_{n\alpha}(k_{z\mu}) + k_y^2 \frac{\hbar^2}{2m_e} \tilde{\chi}_{n\alpha}(k_{z\mu}) + k_{z\mu}^2 \frac{\hbar^2}{2m_e} \tilde{\chi}_{n\alpha}(k_{z\mu}) \\
&+ k_{x,\alpha} \frac{\hbar}{m_e} \sum_m p_{nm}^x(x_\mathrm{c}) \tilde{\chi}_{m\alpha}(k_{z\mu}) + k_y \frac{\hbar}{m_e} \sum_m p_{nm}^y(x_\mathrm{c}) \tilde{\chi}_{m\alpha}(k_{z\mu}) \\
&\pm k_{z\mu} \frac{i\hbar}{m_e} \sum_m p_{nm}^z(x_\mathrm{c}) \tilde{\chi}_{m\alpha}(k_{z\mu}) + \sum_m H_{nm}(x_\mathrm{c}) \tilde{\chi}_{m\alpha}(k_{z\mu}) \\
&+ \sum_{\mu'=0}^{+\infty} \left[\tilde{V}_\mathrm{e}(x_\mathrm{c}, k_{z\mu'} - k_{z\mu}) \mp \tilde{V}_\mathrm{e}(x_\mathrm{c}, k_{z\mu'} + k_{z\mu}) \right] \tilde{\chi}_{n\alpha}(k_{z\mu'}) \\
&= E \tilde{\chi}_{n\alpha}(k_{z\mu})
\end{aligned}
\tag{5.9}
$$

where again the top (bottom) sign corresponds to the odd (even) EFs as in Eq. (5.8). Having determined the spectral EF systems of both the active area and in the contacts, the next step is to discretize the position domain and solve for the EFs.

5.3.4 Discrete spectral transformations

In an actual implementation, the position domain of the spectral decomposition is not continuous as in Eqs. (5.3) and (5.4), but discretized over a spatial device mesh. In order to implement the spectral approach of the previous section, we therefore need a discrete transformation that corresponds to the continuum decompositions of Eqs. (5.3) and (5.4). This transformation converts a finite discrete vector in the position domain to a finite vector in the spectral domain that contains the spectral components. Since Eqs. (5.3) and (5.4) are effectively a continuous sine and cosine transform, the obvious choice for the discrete transforms is a discrete sine transform (DST) and a discrete cosine transform (DCT) respectively.

Several types of DST and DCT exist, depending on the periodic extension of the discrete vector in the position domain beyond the domain boundaries. The continuous spectral decompositions in Eqs. (5.3) and (5.4) effectively extend the domain of the EF periodically beyond the device domain in z: the sine and cosine functions can be evaluated at any z-value. In the continuous case, there are only two types of extensions: odd for the sine expansion and even for the cosine expansion. For a finite discrete EF vector, on the other hand, there are more possibilities. At each endpoint, the extension can be even or odd, referred to in

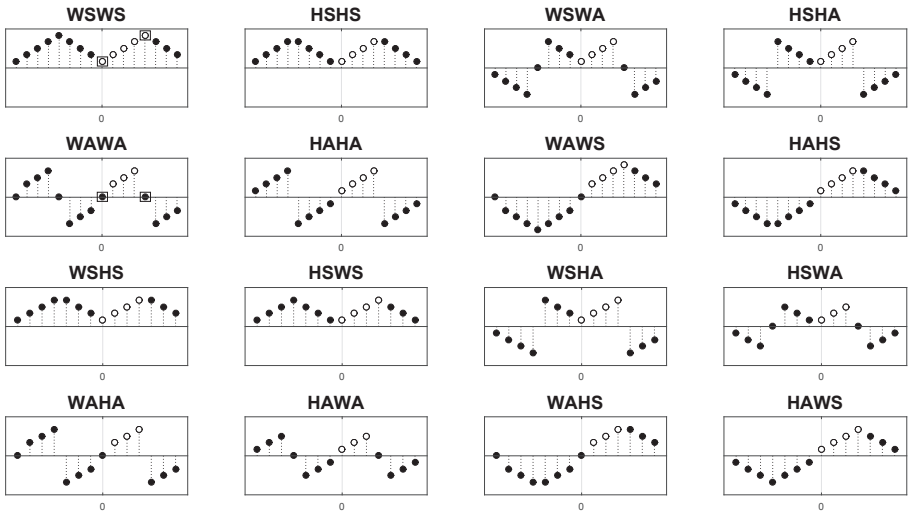

Figure 5.4: Types of (anti)symmetric extension of a discrete vector. Each type is denoted with a four letter acronym. Two letters for each endpoint: W(H) stands for whole(half)-sample symmetry, S(A) for (anti)symmetric. The open symbols are the representative elements, based on which the full extended periodic sequence can be generated if the symmetry of the extension is known. For the extension types that correspond to the chosen DST/DCT transformation types, the elements corresponding to the device boundary points have been indicated with a square box. This figure is based on Fig. 2 of Martucci [151].

the discrete case as respectively "symmetric" and "antisymmetric" extension. Each point of symmetry can also either be an element of the vector ("whole sample symmetry") or it can lie between two points ("half sample symmetry"). Given two endpoints, this results in a total of sixteen types of extensions (see Fig. 5.4). The extension types with an antisymmetric (symmetric) extension at the left endpoint (at $z = 0$) correspond to a type of DST (DCT). The different types of DST/DCT are denoted with Roman numerals, e.g. DST-I, DST-II... A full listing can be found in [151].

It is clear that the choice of DST/DCT and associated extension type determines the exact boundary conditions at the device boundary. For our application, we prefer WAWA extension for the odd EFs: it enforces a zero boundary value (Dirichlet boundary condition) on the boundary mesh points. This extension type corresponds to a DST-I. For the even EFs and the potential energy, on the other hand, we would like to allow for any boundary value. This corresponds to the extension types with symmetric extension at both endpoints, leaving us

with four possible extensions to choose from.

Besides retaining the desired boundary conditions, the chosen combination of DST/DCT should transform a multiplication of two vectors in the position domain in the correct convolution relation in the spectral domain. In our application, this is required for the potential energy term in the EF system, which contains a multiplication in the position domain of the potential energy function and the EF. For the odd EFs, this term is transformed by a combination of a DCT for the potential energy vector and a DST for the EF vector. The transformation of this multiplication with these two different transformation types should in the spectral domain result in the expression as derived for the continuum decomposition in Eq. (5.8), which is a combination of a convolution and an autocorrelation.

Martucci has derived multiplication-convolution properties for all combinations of DST and DCT types [151]. He showed that because of differences in symmetry, not all combinations of DST and DCT types are "compatible" in the sense that a multiplication-convolution relation exists. The expression for the convolution is also different depending on the combination of DST/DCT types. For our application, it turns out that if we combine the DST-I that we selected for the odd EFs with a DCT-I for the even EFs and potential energy function, we obtain both the correct multiplication-convolution relationship and the desired boundary conditions as described above. These transformations correspond to respectively a WAWA and WSWS extension in Fig. 5.4.

The expressions for the chosen transformation types are, for a general vector $\mathbf{x}$:

DCT-I (WSWS):

$$\tilde{\mathbf{X}}(n) = 2\sum_{m=0}^{N-1} \mathbf{x}(m)\cos\left(\frac{nm\pi}{N-1}\right)\mathbf{c}(m) \qquad n = 0..N-1 \tag{5.10}$$

where N is the number of elements in the vector $\mathbf{x}$ (this definition differs from that used by Martucci [151], see also Fig. 2 in that reference), $\mathbf{c}(m) = 1/2$ for $m = \{0, N-1\}$ and 1 otherwise. The inverse transformation is the same expression as Eq. (5.10), multiplied with a factor $1/(2N-2)$.

DST-I (WAWA):

$$\tilde{\mathbf{X}}(n) = 2\sum_{m=0}^{N-1} \mathbf{x}(m)\sin\left(\frac{nm\pi}{N-1}\right) \qquad n = 0..N-1 \tag{5.11}$$

where the inverse transformation is again the same expression as Eq. (5.11), multiplied with a factor $1/(2N-2)$. Note that our indexing convention includes the zero components of the DST: if n or m are zero or $N-1$, the sine factor is zero. We have chosen this convention such that N is the same in the DCT and the DST.

The definitions of DCT-I and DST-I are typically given in literature for vectors with the same size of the original and transformed vector. In our application, however, we discard spectral components which are close to zero after the transformation, such that the number of elements in the spectral domain, N_{k_z}, is generally smaller than the number of elements in the position domain N_z. If an inverse transformation is required, the vector in the spectral domain can be padded with $(N_z - N_{k_z})$ zeros before transformation with Eqs. (5.10) and (5.11).

We now illustrate the indexing, boundary conditions and multiplication-convolution property of the chosen DST/DCT combination with a simple example.

5.3.5 Example

We consider the multiplication of two vectors in the position domain: a potential vector $\mathbf{V}$ and an odd EF vector $\mathbf{F}^{(o)}$. The goal is to convert the multiplication of these two vectors in the position domain to a convolution in the spectral domain as we did for the continuous case in Section 5.3.3 and Appendix B, while maintaining the correct boundary conditions. If the discrete transform types chosen in the previous section are correct, we should expect an expression that is the discrete analog to the potential term of Eq. (5.8).

$\mathbf{V}$ and $\mathbf{F}^{(o)}$ are vectors defined on a single z-slice in the position domain:

$$\mathbf{V} = \begin{bmatrix} V_0 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \end{bmatrix}, \quad \mathbf{F}^{(o)} = \begin{bmatrix} 0 \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ 0 \end{bmatrix} \tag{5.12}$$

where zero Dirichlet boundary conditions are imposed on $\mathbf{F}^{(o)}$, while $\mathbf{V}$ can have any value at the boundary points. To convert the product of these two vectors to the spectral domain, we take the DCT-I of $\mathbf{V}$ and the DST-I of $\mathbf{F}^{(o)}$, as discussed in Section 5.3.4. This corresponds to the following periodic

extensions:

$$\mathbf{V}_{\text{ext}} = \begin{bmatrix} V_0 \\ V_1 \\ V_2 \\ V_3 \\ V_4 \\ V_5 \\ V_4 \\ V_3 \\ V_2 \\ V_1 \end{bmatrix}, \quad \mathbf{F}^{(o)}_{\text{ext}} = \begin{bmatrix} 0 \\ F_1 \\ F_2 \\ F_3 \\ F_4 \\ 0 \\ -F_4 \\ -F_3 \\ -F_2 \\ -F_1 \end{bmatrix} \tag{5.13}$$

where one full period of the infinite periodic sequence is shown and where the type of the extension is WSWS for $\mathbf{V}_{\text{ext}}$ and WAWA for $\mathbf{F}^{(o)}_{\text{ext}}$. Using Eqs. (5.10) and (5.11) (with $N = 6$), we obtain the transformed vectors in the spectral domain:

$$\tilde{\mathbf{V}} = \begin{bmatrix} \tilde{V}_0 \\ \tilde{V}_1 \\ \tilde{V}_2 \\ \tilde{V}_3 \\ \tilde{V}_4 \\ \tilde{V}_5 \end{bmatrix}, \quad \tilde{\mathbf{F}}^{(o)} = \begin{bmatrix} 0 \\ \tilde{F}_1 \\ \tilde{F}_2 \\ \tilde{F}_3 \\ \tilde{F}_4 \\ 0 \end{bmatrix} \tag{5.14}$$

Martucci showed that for a DST/DCT combination of type I a multiplication of the vectors in the position domain corresponds to a circular convolution of the extended vectors in the spectral domain [151]. We therefore first write a periodic extension of the transformed vectors, whereby the extension of the transformed vectors needs to be identical to the extension of the position domain vectors:

$$\tilde{\mathbf{V}}_{\text{ext}} = \begin{bmatrix} \tilde{V}_0 \\ \tilde{V}_1 \\ \tilde{V}_2 \\ \tilde{V}_3 \\ \tilde{V}_4 \\ \tilde{V}_5 \\ \tilde{V}_4 \\ \tilde{V}_3 \\ \tilde{V}_2 \\ \tilde{V}_1 \end{bmatrix}, \quad \tilde{\mathbf{F}}^{(o)}_{\text{ext}} = \begin{bmatrix} 0 \\ \tilde{F}_1 \\ \tilde{F}_2 \\ \tilde{F}_3 \\ \tilde{F}_4 \\ 0 \\ -\tilde{F}_4 \\ -\tilde{F}_3 \\ -\tilde{F}_2 \\ -\tilde{F}_1 \end{bmatrix} \tag{5.15}$$

The circular convolution is defined as:

$$\tilde{\mathbf{R}}_{\text{ext}}(n) = \sum_{k=0}^{n} \tilde{\mathbf{V}}_{\text{ext}}(k)\tilde{\mathbf{F}}^{(o)}_{\text{ext}}(n-k) + \sum_{k=n+1}^{M-1} \tilde{\mathbf{V}}_{\text{ext}}(k)\tilde{\mathbf{F}}^{(o)}_{\text{ext}}(n-k+M) \tag{5.16}$$

with $M = 2N - 2 = 10$ the number of elements in the extended vectors and where we use zero-based indexing. If we reverse the ordering of the terms in the summations, we can also express Eq. (5.16) as a matrix multiplication with a so-called "Toeplitz matrix":

$$\tilde{\mathbf{R}}_{\text{ext}} = \begin{bmatrix} \tilde{R}_0 \\ \tilde{R}_1 \\ \tilde{R}_2 \\ \tilde{R}_3 \\ \tilde{R}_4 \\ \tilde{R}_5 \\ \tilde{R}_6 \\ \tilde{R}_7 \\ \tilde{R}_8 \\ \tilde{R}_9 \end{bmatrix} = \begin{bmatrix} \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 \\ \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 \\ \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 \\ \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 \\ \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 \\ \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 \\ \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 \\ \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 & \tilde{V}_2 \\ \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 & \tilde{V}_1 \\ \tilde{V}_1 & \tilde{V}_2 & \tilde{V}_3 & \tilde{V}_4 & \tilde{V}_5 & \tilde{V}_4 & \tilde{V}_3 & \tilde{V}_2 & \tilde{V}_1 & \tilde{V}_0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \tilde{F}_1 \\ \tilde{F}_2 \\ \tilde{F}_3 \\ \tilde{F}_4 \\ 0 \\ -\tilde{F}_4 \\ -\tilde{F}_3 \\ -\tilde{F}_2 \\ -\tilde{F}_1 \end{bmatrix} \tag{5.17}$$

We can rewrite this expression as a matrix multiplication of the size of $\tilde{\mathbf{V}}$ and $\tilde{\mathbf{F}}^{(o)}$ before extension (6×1):

$$\tilde{\mathbf{R}} = \begin{bmatrix} \tilde{R}_0 \\ \tilde{R}_1 \\ \tilde{R}_2 \\ \tilde{R}_3 \\ \tilde{R}_4 \\ \tilde{R}_5 \end{bmatrix} = \begin{bmatrix} \tilde{V}_0 & 0 & 0 & 0 & 0 & 0 \\ \tilde{V}_1 & \tilde{V}_0 - \tilde{V}_2 & \tilde{V}_1 - \tilde{V}_3 & \tilde{V}_2 - \tilde{V}_4 & \tilde{V}_3 - \tilde{V}_5 & \tilde{V}_4 \\ \tilde{V}_2 & \tilde{V}_1 - \tilde{V}_3 & \tilde{V}_0 - \tilde{V}_4 & \tilde{V}_1 - \tilde{V}_5 & \tilde{V}_2 - \tilde{V}_4 & \tilde{V}_3 \\ \tilde{V}_3 & \tilde{V}_2 - \tilde{V}_4 & \tilde{V}_1 - \tilde{V}_5 & \tilde{V}_0 - \tilde{V}_4 & \tilde{V}_1 - \tilde{V}_3 & \tilde{V}_2 \\ \tilde{V}_4 & \tilde{V}_3 - \tilde{V}_5 & \tilde{V}_2 - \tilde{V}_4 & \tilde{V}_1 - \tilde{V}_3 & \tilde{V}_0 - \tilde{V}_2 & \tilde{V}_1 \\ \tilde{V}_5 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ \tilde{F}_1 \\ \tilde{F}_2 \\ \tilde{F}_3 \\ \tilde{F}_4 \\ 0 \end{bmatrix} \tag{5.18}$$

or alternatively:

$$\tilde{\mathbf{R}}(n) = \sum_{n'=0}^{N-1} \left[\left(\tilde{\mathbf{V}}(n') - \tilde{\mathbf{V}}(n) \right) - \left(\tilde{\mathbf{V}}(n') + \tilde{\mathbf{V}}(n) \right) \right] \tilde{\mathbf{F}}(n') \tag{5.19}$$

which corresponds to the potential energy term for the odd EFs in Eq. (5.8), with the indices n, n' corresponding to μ, μ'. The derivation for the even EFs is completely analogous, except that all the minus signs in Eq. (5.18) become a plus sign. Note that in a numerical implementation the first and last columns and rows of the matrix in Eq. (5.18) can be removed, since these are either zero or multiplied with a zero element of $\tilde{\mathbf{F}}^{(o)}$.

If desired, an inverse DST-I of the result vector $\tilde{\mathbf{R}}$ in the spectral domain gives us the result of the multiplication of the original vectors $\mathbf{V}$ and $\mathbf{F}^{(o)}$ in the position domain [151].

We have shown in this section that we can combine different transforms to apply different boundary conditions to vectors in position space and that it

is possible to express a multiplication of these vectors as a convolution or a matrix multiplication in the transformed space. The full spectral discretized EF system is presented in Appendix A.3.

5.3.6 Spectral band structure

Repeating the band structure calculation of Fig. 5.3 with the discretized spectral version of the contact eigenvalue problem in Eq. (5.9) results in a band structure without purely real zone-center spurious solutions (see Fig. 5.5). The subbands belonging to the different bands are clearly discernible, with the imaginary band structure showing the various couplings between them. In contrast to the bulk band structure of Fig. 5.2, most confined (sub)bands are no longer purely associated to a single basis state, but are a mixture of various basis states.

The absence of zone-center spurious solutions results from the limitation of the sampling in the k_z-direction, which comes from the frequency selection property of spectral methods discussed in Section 5.3.1. The number of spectral components N_{k_z} in Eq. (5.9) determines the amount of slices taken in the k_z-direction (see Fig. 4.1), and hence the number of subbands considered (compare Fig. 5.5(a) to Fig. 5.5(b)). While in a finite difference approach the full band structure is sampled, including the spurious upwards curvature of the valence band, a spectral method with limited N_{k_z} samples only the physical region of the band structure at low k_z-values. Therefore, no out-of-zone spurious solutions are mapped back onto the zone center. Thanks to the exponential accuracy of the spectral approach, N_{k_z} in any practical TFET application is small enough to limit the sampling to the physical region of the band structure, while being accurate enough to represent the EFs.

Although the real zone-center spurious solutions are gone, we do still see complex solutions inside of the first Brillouin zone. We can divide them into two categories, indicated in Fig. 5.5 as c_1 and c_2. c_1-solutions have a large real k_x-value and are a result of the limited EF basis set. We handle them like the out-of-zone spurious solutions (see Section 4.2.3). c_2-solutions, however, couple the valence band and the first conduction band. The imaginary part of these solutions are analogous to what is found in a confined two-band model. Here, however, they also have a significant real part, seemingly closing up the bandgap. The explanation for this real part can be found in the shape of the valence band when zoomed in (see Fig. 5.6): the valence band subbands do not reach their maximum at the Γ-point, but at a point where k_x is not zero. This is a result of the non-isotropic shape of the valence band, which is projected onto the $k_z = 0$-plane when confined in the z-direction. In a semiclassical picture, this means that when a carrier is traveling in the transport direction x and

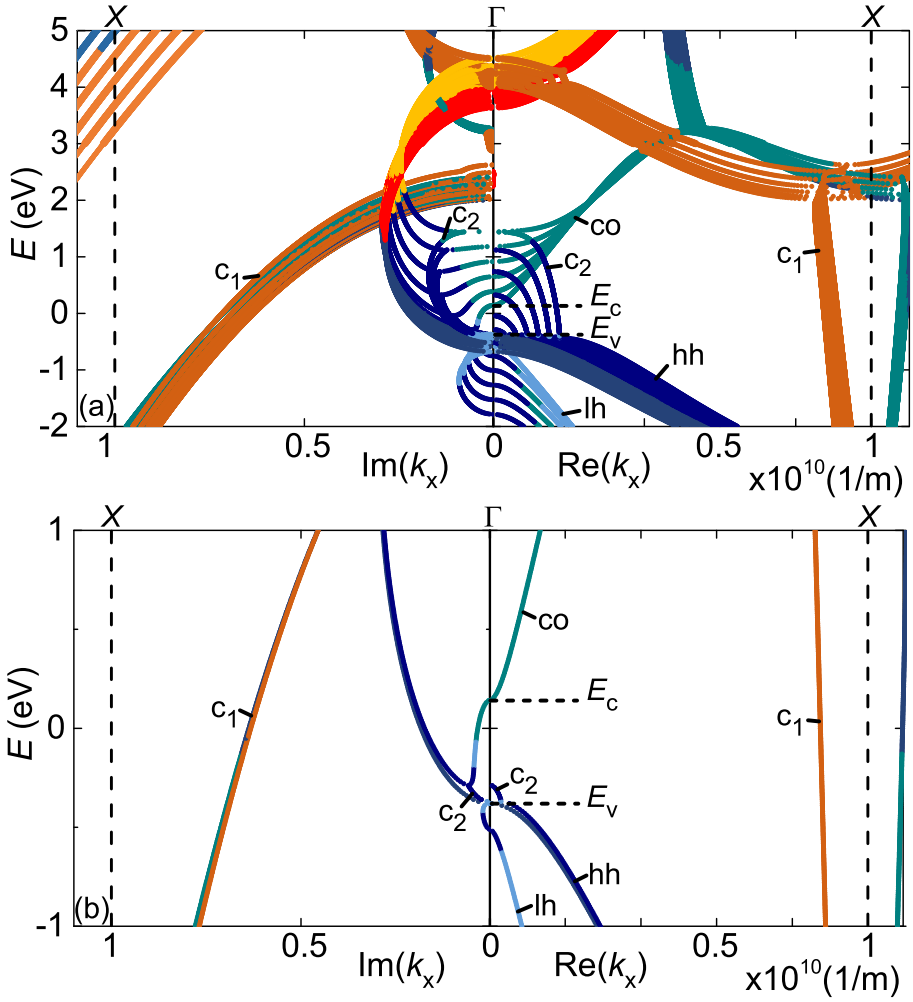

Figure 5.5: Fifteen-band confined real (right) and imaginary (left) band structure of a 10 nm InAs slab in the [100] transport direction, calculated using a spectral method in the confined direction with (a) 6 and (b) 1 spectral components. k_y is zero. The dashed lines indicate the edges of the first Brillouin zone. The solutions are colored based on the dominant component in the EF vector, corresponding to the maximum band contribution to each solution. Indicated are the conduction (co), light hole (lh) and heavy hole (hh) bands and two types of complex solutions (c_1 and c_2). The fifteen-band parameters are taken from Radhia *et al.* [58] and are listed in Appendix C.2.

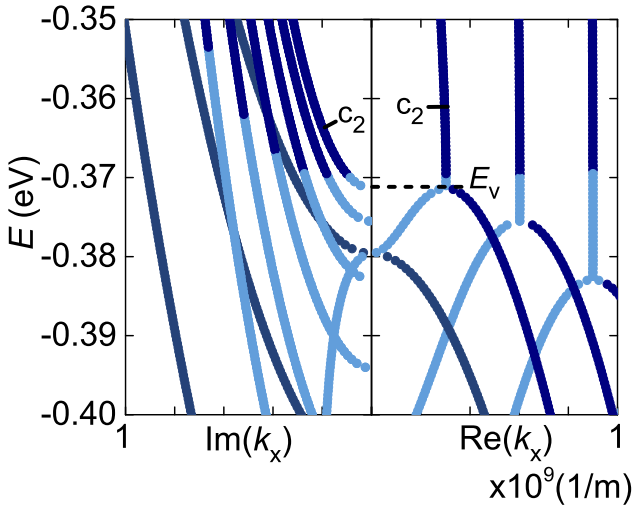

Figure 5.6: Zoom around the valence band edge of the confined real (right) and imaginary (left) band structure in Fig. 5.5(a).

reaches the band edge, its k_x-value turns from purely real to complex, instead of purely imaginary. The carrier then follows a path through the complex space from one band edge to the other. Along this path, the real part of k_x does not influence the probability density, which is decaying inside the forbidden region due to the imaginary part. Therefore, we handle the c_2-solutions like the purely imaginary states in the two-band model (see Section 4.2.3): we do not inject them, but keep them in the boundary conditions for reflection or transmission.

In summary, with the spectral approach, the zone-center spurious solutions are now either removed or handled in a consistent manner. The band structure model is therefore now ready for device simulations. However, the band structures shown in this and previous sections have been exclusively of InAs. To enable a wider variety of material systems for TFET simulations, we first develop a procedure to determine fifteen-band parameters in Section 5.4, before discussing the device simulations in the subsequent sections.

5.4 Parameter calibration

Fifteen-band parameter sets are more difficult to obtain than sets for the two-band model, which are only determined by E_G, p and the orthogonal effective masses, as discussed in Section 4.3. Firstly, III-V parameter sets for models with more than eight bands are available for only a few materials [58, 152, 153]. Secondly, the available sets typically lack commutativity of the momentum operator, which is an essential property required by the heterostructure basis transform discussed in Section 3.2.

In an infinite band model, commutativity of the momentum operator in different directions is ensured as a fundamental quantum mechanical property [154]. The corresponding infinite-dimensional momentum matrices have the following commutativity property:

$$\begin{aligned}
[\mathbf{P}_x, \mathbf{P}_y] &= \mathbf{P}_x\mathbf{P}_y - \mathbf{P}_y\mathbf{P}_x = 0 \\
[\mathbf{P}_x, \mathbf{P}_z] &= \mathbf{P}_x\mathbf{P}_z - \mathbf{P}_z\mathbf{P}_x = 0 \\
[\mathbf{P}_y, \mathbf{P}_z] &= \mathbf{P}_y\mathbf{P}_z - \mathbf{P}_z\mathbf{P}_y = 0
\end{aligned} \tag{5.20}$$

with $\mathbf{P}_x$, $\mathbf{P}_y$, $\mathbf{P}_z$ the momentum matrices in the x-, y- and z-direction, containing the interband momentum matrix elements p^x_{nm}, p^y_{nm} and p^z_{nm} respectively. This commutativity property is equivalent to stating that the momentum matrices have a common eigenvalue decomposition, which is used in Eqs. (3.11) and (3.12) to construct the heterostructure basis function transformation. This commutativity is preserved in an infinite band model description, but is at risk in a finite band model.

When fitting the parameters of a finite band model, the interband momentum matrix elements are varied to match the resulting band structure to a certain reference, e.g. experimental data at symmetry points or an ab-initio band structure. In this process, the commutativity property of Eq. (5.20) is typically lost. For applications that only use the band structure to extract effective masses, this is not an issue. For our application, however, the commutativity property is required to construct the heterostructure transformation matrices.

We therefore develop a procedure which allows for the fitting of multi-band parameter sets, while retaining the commutativity of the momentum matrices. A flowchart is shown in Fig. 5.7. The core of the procedure is a constrained optimization algorithm. This algorithm minimizes a scoring function, which compares the band structure calculated with the parameters of the current iteration with given reference values. These reference values are typically experimental values of effective masses and energy gaps at high symmetry points

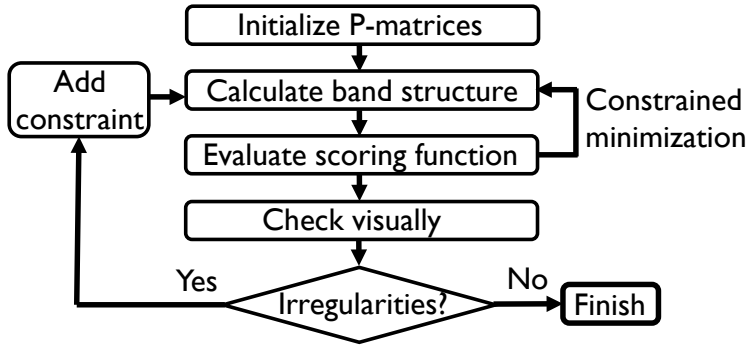

Figure 5.7: The procedure used to obtain multi-band parameters that retain commutativity of the momentum matrices.

in the band structure. In more detail, the scoring function ϕ is constructed as a weighted sum of differences:

$$\phi = w_1\Delta m^*_{\mathrm{el}} + w_2\Delta m^*_{\mathrm{hh}} + w_3\Delta m^*_{\mathrm{lh}} + w_4\Delta m^*_X + w_5\Delta m^*_L + w_6\Delta E_X + w_7\Delta E_L \tag{5.21}$$

where Δm^*_{el}, Δm^*_{hh}, Δm^*_{lh}, Δm^*_X and Δm^*_L are the differences in curvature effective mass at the Γ-point for the conduction band, the heavy hole band and the light hole band, and for the conduction band at the X- and L-point respectively. ΔE_X and ΔE_L are the differences in energy gaps between conduction and valence band at the X- and L-points. The band edges at the Γ-point are fixed parameters, and are not varied during the fitting procedure. w_1 to w_7 are weights that can put more or less importance on the corresponding differences, which correspond to specific features of the band structure. If desired, terms can be added or removed from the scoring function, based on the availability of reliable reference values. The optimization algorithm searches for the set of interband momentum matrix elements that minimizes the scoring function, within the constraint of Eq. (5.20), namely that the momentum matrices have to commute. The optimization algorithm is supplied by the NLopt package [155], which solves multivariable non-linear constrained optimization problems. A visual check is carried out when the algorithm finishes, to make sure the band structure does not show any irregularities, e.g. a band that curves into the bandgap. Additional constraints can be added to avoid such irregularities, after which the optimization algorithm is run again.

The parameter sets resulting from this fitting procedure are used throughout the remainder of the thesis and are listed in Appendix C.2.

5.5 N-channel pocketed pointTFET vs lineTFET in direct bandgap III-V materials

With our implemented spectral EF formalism and a procedure to obtain fifteen-band material parameters, we now address one of the main goals of this thesis: the comparison of the pocketed lineTFET with the pocketed pointTFET in a direct bandgap III-V material system. In this section, we make the comparison for homostructure nTFET configurations. We first verify that the lineTFET concept still holds up in a direct bandgap III-V material system (Section 5.5.2). We then compare fifteen-band to two-band simulations and check whether we actually need fifteen-band simulations to do the pointTFET-lineTFET comparison (Section 5.5.3). Finally, we assess whether an optimized pocketed lineTFET or pointTFET is most promising (Section 5.5.4).

5.5.1 Simulation details

The simulated TFET configurations are shown in Fig. 5.8, with the details in Table 5.1. Doping profiles in these configurations are abrupt and uniform. The material of choice is $In_{0.53}Ga_{0.47}As$, of which the fifteen-band QM parameters can be found in Appendix C.2.1 and the SC material parameters used for the calculation of the electrostatic potentials with SDevice in Appendix D.3.2.

Numerically, we take 25 spectral components for the lineTFET configurations and 6 for the pointTFET. This number is higher for the lineTFET, because the configurations are wider and the electrostatic potential varies more in the confined z-direction. The x-direction is discretized with a finite difference mesh of 1000 points. In the y-direction, which is translationally invariant, we take 10 k_y-points. The energetic tunneling window is discretized adaptively in 40 points, which means the mesh is refined around peaks in the integrand of the current integral in Eq. (3.56). We have checked in all cases that the device current no longer changes when increasing the number of mesh points or spectral components.

To speed up the simulations, we parallelize the execution of the different k_y-points [156]. The calculation of the current for each of these points is "embarrassingly parallel", which means the parallel processes only communicate at the end to collect the partial results and pass them on to the leader process. Because of the low communication overhead, the scaling of the simulation time is essentially linear with the amount of parallel processes.

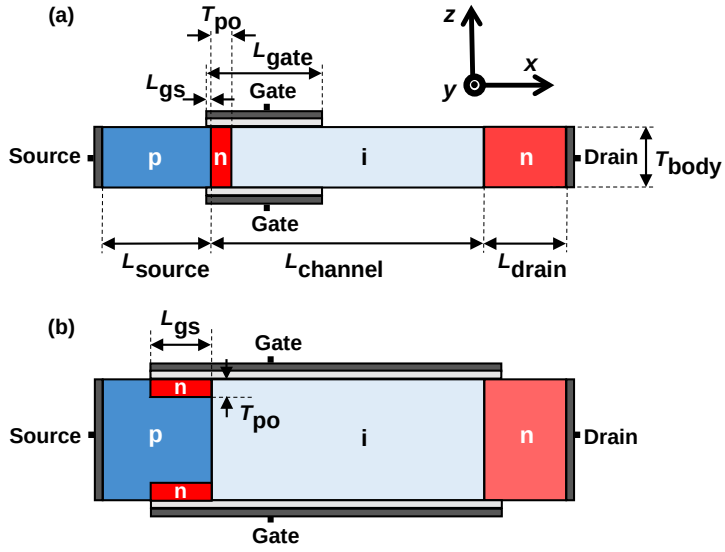

Figure 5.8: Investigated nTFET configurations: (a) pocketed pointTFET and (b) pocketed lineTFET.

Table 5.1: Details of the simulated nTFET configurations (see Fig. 5.8).

	point	line
L_{source} [nm]	20	20
L_{channel} [nm]	50	50
L_{drain} [nm]	15	15
L_{gate} [nm]	22	65
L_{gs} [nm]	1	10
T_{body} [nm]	varying	40
T_{po} [nm]	varying	varying
Doping source [cm^{-3}]	$1\text{x}10^{20}$	$1\text{x}10^{20}$
Doping pocket [cm^{-3}]	$1\text{x}10^{20}$	$1\text{x}10^{20}$
Doping channel [cm^{-3}]	$1\text{x}10^{10}$	$1\text{x}10^{10}$
Doping drain [cm^{-3}]	$5\text{x}10^{17}$	$5\text{x}10^{17}$
EOT [nm]	0.6	0.6
GWF [eV]	5	5

5.5.2 Viability of direct bandgap III-V lineTFET

We start by simulating a single pocketed direct bandgap III-V lineTFET to assess the effect of the directional change in the tunneling current that occurs right beyond the tunnel junction (see inset of Fig. 5.9). For group IV materials with an indirect bandgap, such as Si, pocketed lineTFETs have shown promise, with QM simulations showing lower SS and higher I_{ON} than pocketed pointTFETs [13] (see also Section 2.8.2). In direct bandgap materials, however, there is no phonon involved in the tunneling transition that can account for the change in k-vector required in the directional change of the current flow. In addition, the effect of FIQC is expected to be more severe in III-V materials, because of the lower effective mass of the conduction band. It is therefore a priori unclear whether the lineTFET concept can hold up in a III-V material system. To assess the impact of the directional change and FIQC, we compare the QM simulations to SC simulations carried out with SDevice. SDevice calculates the tunneling current with a non-local BTBT generation model, similar to the one described in Section 2.5.1 [157]. This model is impacted neither by the

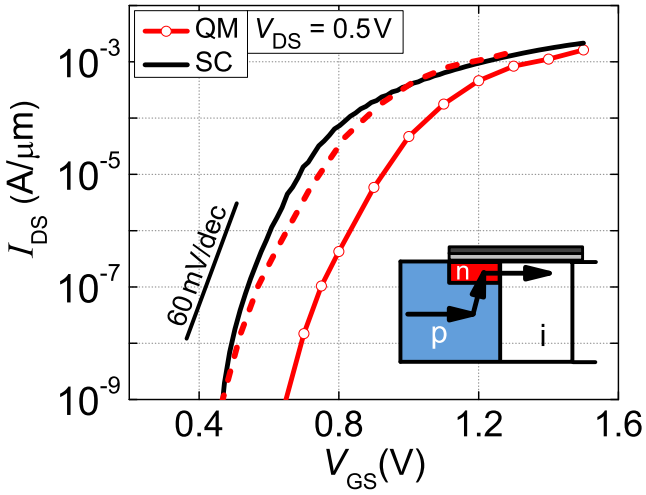

Figure 5.9: Comparison of SC and QM simulated transfer characteristics of a pocketed lineTFET with a T_{po} of 4 nm. The dashed line is the QM curve shifted such that the V_{GS} at which I_{OFF} is 1 nA/μm coincides. Inset: schematic path of the tunneling current at the source-pocket junction. More details on the configuration can be found in Fig. 5.8(b) and Table 5.1.

directional change in current, nor by FIQC.

Fig. 5.9 shows that despite the directional change in current and the presence of FIQC, the SC and QM simulations predict similar lineTFET performance. Both curves predict I_{ON} to be around $3\mathrm{x}10^{2}$ μA/μm for a supply voltage window of 0.5 V and an I_{OFF} defined at 1 nA/μm. I_{60} is around $2\mathrm{x}10^{1}$ nA/μm. This indicates that the electric field in the pocket region has a sufficient component in the direction of the drain to accomplish the directional change in current. As to the FIQC, the main effect is a shift of V_{onset} by about 180 mV, while the SS is only slightly degraded between a V_{GS} of 0.5 V and 1 V (in the shifted version). The conclusion of this comparison is that the lineTFET concept does not suffer dramatic current decreases and hence holds up in a QM analysis for a direct bandgap III-V material system.

5.5.3 Comparison two-band to fifteen-band simulations

Having verified that our fifteen-band implementation is able to simulate lineTFETs, we now compare fifteen-band simulation results to the two-band results of the previous chapter, before starting with the pointTFET-lineTFET comparison in Section 5.5.4. With the full-zone fifteen-band model, we first verify our choice in Section 4.2.2 to take the heavy hole mass as the orthogonal valence band effective mass in the two-band model. Next, we check the discrepancies between the models for more confined structures, both diodes and TFETs. And finally, we assess whether we actually need the fifteen-band model to carry out the pointTFET-lineTFET comparison.

Orthogonal effective mass

Fig. 5.10 compares the two-band transfer characteristics of Fig. 4.3 for the diode with a T_{i} of 9 nm with a fifteen-band simulation of the same configuration. It is clear that the two-band simulation with the heavy hole orthogonal mass corresponds best to the fifteen-band simulation. This confirms the theoretical reasoning followed in Section 4.2.2 that the orthogonal DOS is dominated by the heavy hole band.

Confinement

Fig. 5.11 repeats the comparison between two-band and fifteen-band diode simulations of Fig. 5.10 for varying T_{body}. While the two-band and fifteen-band simulations match well for the wider diodes of 60 nm and 30 nm, the 10 nm

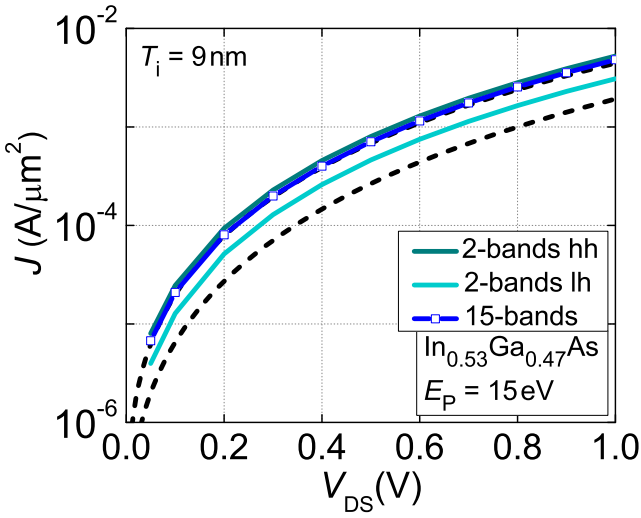

Figure 5.10: Two- and fifteen-band QM simulations of current densities for the p-i-n diode of Fig. 4.3 with a 9 nm T_{i}. The two-band simulations are carried out with a light hole (lh) and heavy hole (hh) m_{v}^*. The dashed curves are SC simulations carried out with the upper and lower values for A_{BTBT} in Table 4.1.

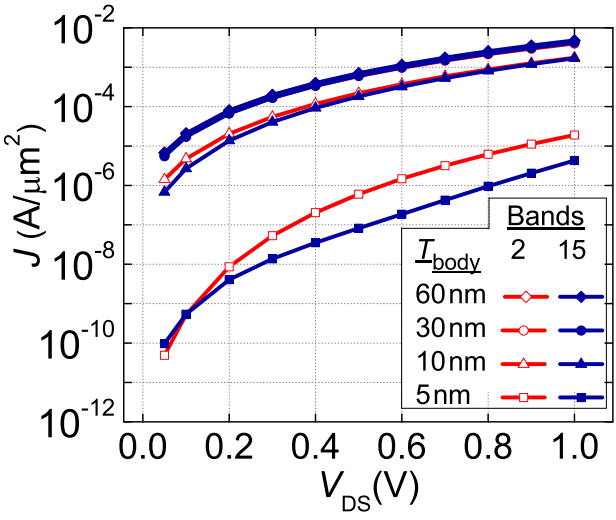

Figure 5.11: Two- and fifteen-band QM simulations of current densities for the p-i-n diode of Fig. 4.3 with varying T_{body}. The electrostatic potential for a given T_{body} is identical for both band models and is constant in the confined direction. The two-band simulations are carried out with a heavy hole m_{v}^*.

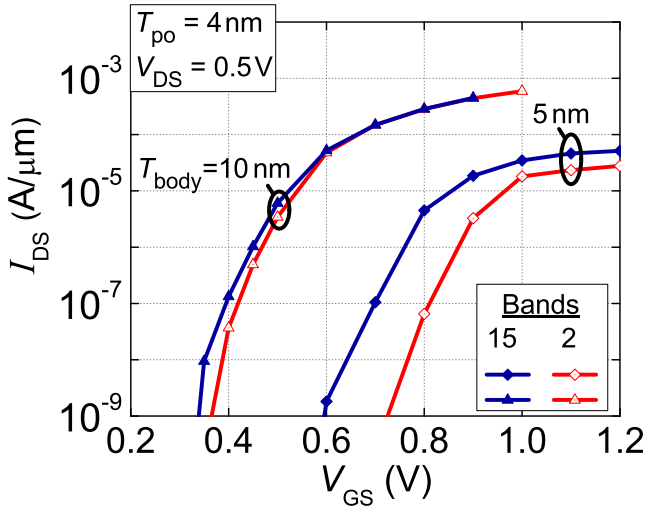

Figure 5.12: Two- and fifteen-band QM transfer characteristics of a pocketed pointTFET with varying T_{body}. The electrostatic potential for a given T_{body} is identical for both band models. More configuration details in Fig. 5.8 and Table 5.1.

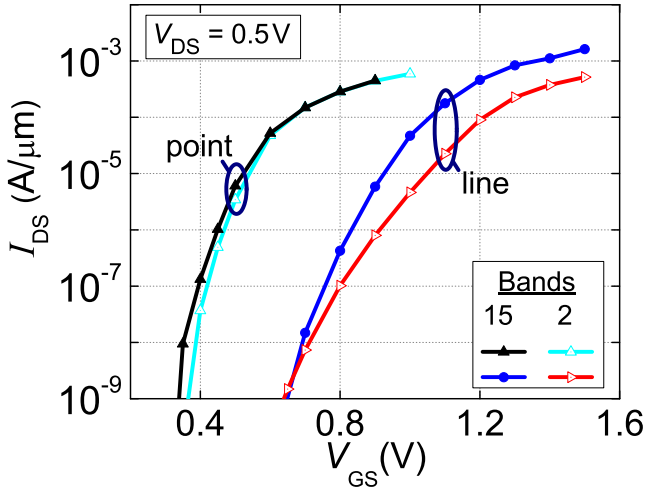

Figure 5.13: Two- and fifteen-band QM transfer characteristics of a pocketed pointTFET and lineTFET, each with a 4 nm T_{po}. T_{body} of the pointTFET is 10 nm. The electrostatic potential of a given configuration is identical for both band models. More configuration details in Fig. 5.8 and Table 5.1.

results start to show a discrepancy for the lower V_{DS}-values. For the most confined structures of 5 nm, the results are strongly diverged.

This behavior stems from the differences in two-band and fifteen-band band structures. The band structures match well around the Γ-point, since they have been fitted to the same experimental effective mass, but they diverge for higher k-values. Away from the Γ-point, the two-band model becomes increasingly inaccurate, while the fifteen-band model captures the full first Brillouin zone. Simulations of more confined structures are more sensitive to the differences in band structure away from the Γ-point, since the k_z-value of the first subband slice (determined by π/T_{body}, see Fig. 4.1) increases as T_{body} decreases. Additionally, the heavy hole and light hole bands in the fifteen-band model split up when confined, since the light hole band is shifted more by the confinement. In the two-band model, no such splitting occurs, because the valence band consists of a single band, which is light in the transport direction and heavy in the orthogonal direction. The splitting reduces the current in the fifteen-band case, because the downward shift of the light hole band moves it partially outside of the tunneling window. The tunneling that remains to the confined heavy hole band is less efficient. At very low V_{DS}, however, we see a cross-over in the current for the most confined configuration. This is because the tunneling window closes faster in the two-band case as the conduction band at the drain side is confined more strongly than in the fifteen-band case.

Fig. 5.12 confirms the trend for a pocketed pointTFET. In line with the diode results, a good qualitative match is obtained for a T_{body} of 10 nm, while the results strongly diverge for a more confined T_{body} of 5 nm. Here, however, we see a higher current for the fifteen-band case.

The results of this comparison validate the two-band diode and TFET results in Section 4.4 down to a T_{body} of 10 nm. Even below 10 nm, however, the same qualitative trend of a counteracting effect between gate control and SIQC-induced bandgap widening is still observed in the fifteen-band model.

Configuration

Next to confined structures, Fig. 5.13 shows that the two-band and fifteen-band transfer characteristics also diverge significantly for a lineTFET configuration. As discussed in Section 4.2.1, this is the result of a lack of coupling, and hence BTBT, in the orthogonal directions in the two-band model. Although BTBT in the transport direction is sufficient for the simulation of devices where the electric field is dominant in this direction, like the pocketed pointTFET, it is inadequate for devices where the electric field is oriented in the perpendicular z-direction, as is the case for a lineTFET. The two-band lineTFET simulations

underestimate the current, because they only capture the lateral tunneling component through a parasitic p-n-i-n pocketed pointTFET with the tunnel junction located at the lateral edge of the source pocket. From these results, it is clear that a fifteen-band model is indeed necessary for the pointTFET-lineTFET comparison.

5.5.4 PointTFET versus lineTFET

For the pointTFET-lineTFET comparison, we optimize the pocketed pointTFET for T_{po} and T_{body}, and the pocketed lineTFET for T_{po}, both for a high source doping (see Table 5.1). The lineTFET is not optimized for T_{body}, since above a critical thickness, lineTFET performance is independent of T_{body} (see Section 2.9.1). For the pointTFET, Fig. 5.14 shows that I_{ON} improves with increasing T_{po} and reaches a plateau above 4 nm. I_{60} reaches a maximum around a T_{po} of 3-4 nm. For larger T_{po}, I_{60} degrades as the structure essentially becomes a p-n diode in series with a n-i-n MOSFET (as discussed in Section 2.8.1). This degradation is more pronounced for the 20 nm T_{body} configuration. I_{60} is in general higher for the 10 nm T_{body} configurations thanks to better gate control throughout the device body. This is in line with our findings in Chapter 4. The best performance is obtained for a T_{po} of 4 nm and a T_{body} of 10 nm, with an I_{ON} of 5x10^2 μA/μm for a V_{DD} of 0.5 V and I_{OFF} of 1 nA/μm. I_{60} is 1 μA/μm. For the lineTFET, performance improves with T_{po}, reaching an I_{ON} of 4x10^2 μA/μm and an I_{60} of 1 μA/μm for a T_{po} of 6 nm and at the same V_{DD} and I_{OFF} as the pointTFET.

Fig. 5.15 compares the transfer characteristics of the best configurations of Fig. 5.14, showing that similar performance can be obtained for the point and lineTFET. This is in line with existing literature on the pointTFET-lineTFET comparison for group IV materials, where it was found that the advantage of the lineTFET over the pointTFET decreases in going to materials with a smaller bandgap [13]. So although a lineTFET configuration is clearly superior for a material with a relatively wide bandgap, such as Si, it is comparable in performance for III-V materials with a smaller bandgap, such as $In_{0.53}Ga_{0.47}As$. This finding actually tips the scale in favor of the pointTFET, which is easier to fabricate and does not require a minimal T_{body} like the lineTFET.

The performance of the most optimized pointTFET and lineTFET presented in this section shows promise to reach the desired specifications discussed in Section 2.6. I_{ON} is generally larger than 10^2 μA/μm at a V_{DD} of 0.5 V. I_{60}, however, still falls short of the 10 μA/μm target. We will therefore discuss the optimization of heterostructure pointTFETs in Section 5.7.

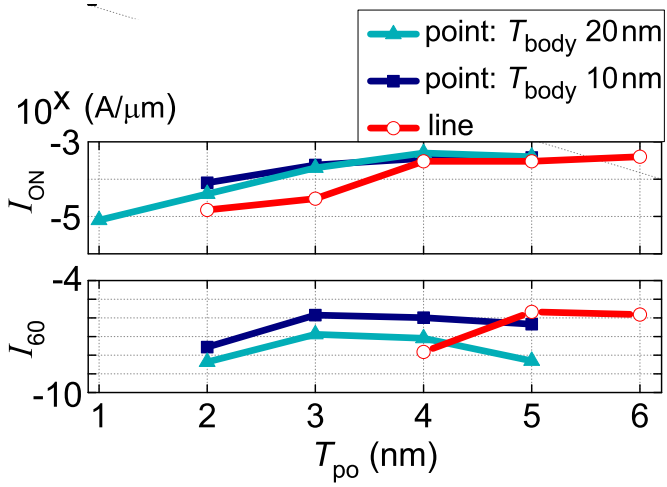

Figure 5.14: QM optimization of I_{ON} and I_{60} for pocketed pointTFET configurations by variation of T_{body} and T_{po} and a pocketed lineTFET by variation of T_{po}. I_{ON} is defined at a V_{GS} which is 0.5 V beyond V_{onset} for an I_{OFF} of 1 nA/μm. More details on the configurations can be found in Fig. 5.8 and Table 5.1.

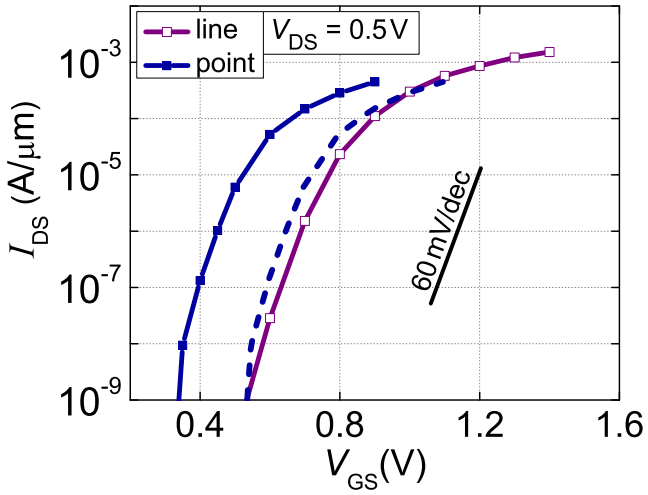

Figure 5.15: Comparison of QM simulated transfer characteristics of pocketed lineTFET and pointTFET, with optimized T_{po} of 6 nm and 4 nm respectively. The pointTFET has a T_{body} of 10 nm. The dashed line is the QM curve shifted such that the V_{GS} at which I_{OFF} is 1 nA/μm coincides. More details on the configurations can be found in Fig. 5.8 and Table 5.1.

Table 5.2: Details of simulated pTFET configurations.

	point	line
L_{source} [nm]	20	20
L_{channel} [nm]	50	50
L_{drain} [nm]	15	15
L_{gate} [nm]	22	65
L_{gs} [nm]	1	10
L_{ud} [nm]	5	/
T_{body}[nm]	varying	40
T_{po} [nm]	varying	varying
Doping n-source [cm^{-3}]	$5\text{x}10^{19}$	$5\text{x}10^{19}$
Doping n^--source [cm^{-3}]	$1\text{x}10^{18}$	/
Doping hetero-source [cm^{-3}]	$5\text{x}10^{17}$	/
Doping n^+-pocket [cm^{-3}]	$5\text{x}10^{19}$	/
Doping p^+-pocket [cm^{-3}]	$5\text{x}10^{19}$	$5\text{x}10^{19}$
Doping channel [cm^{-3}]	$1\text{x}10^{10}$	$1\text{x}10^{10}$
Doping drain [cm^{-3}]	$5\text{x}10^{17}$	$1\text{x}10^{17}$
EOT [nm]	0.6	0.6
GWF [eV]	5	5

5.6 Improved pTFET source design

Although the III-V nTFET configurations in Section 5.5 show promising performance, a complementary technology requires a similarly well performing pTFET. From literature, however, we know this is generally not the case for III-V pTFETs [26, 158, 159]. In this section, we therefore focus on elucidating the origins of the poor pTFET performance (Section 5.6.2) and propose design modifications to bring pTFET performance up to the level of nTFET (Section 5.6.3).

5.6.1 Simulation details

The simulated configurations are similar to the nTFET case and are shown in Fig. 5.16, with details in Table 5.2. We also retain the same material ($\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$) and numerical discretization (see Section 5.5.1) as in the nTFET case.

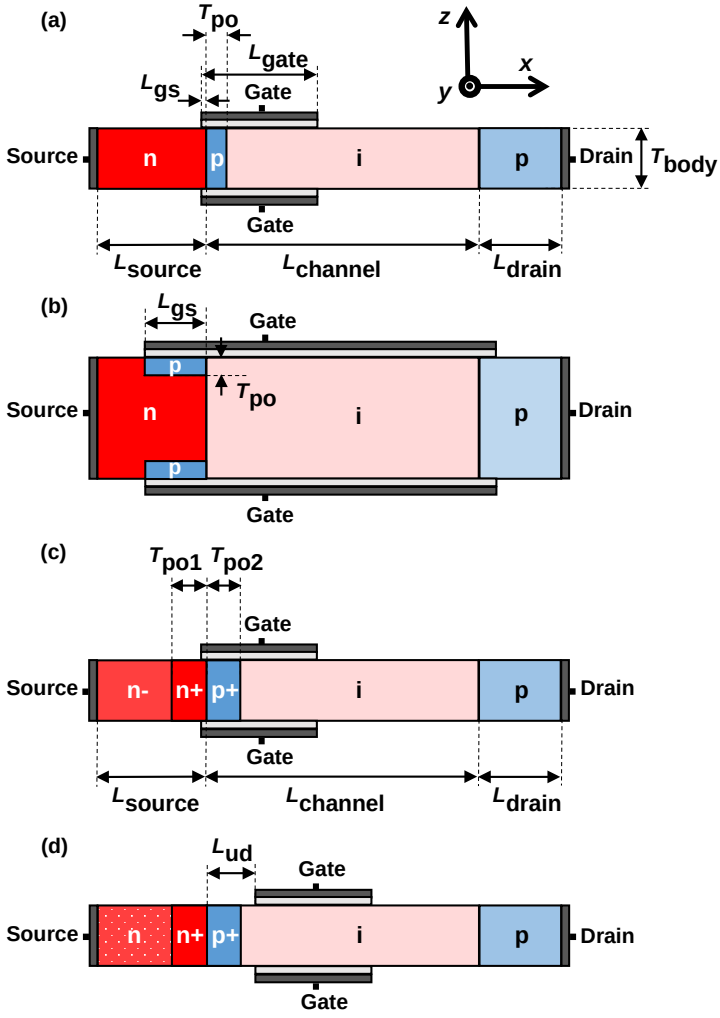

Figure 5.16: Investigated pTFET configurations: (a) pocketed pointTFET, (b) pocketed lineTFET, (c) pocketed pointTFET with improved source design and (d) pocketed pointTFET with improved source design with a heterostructure.

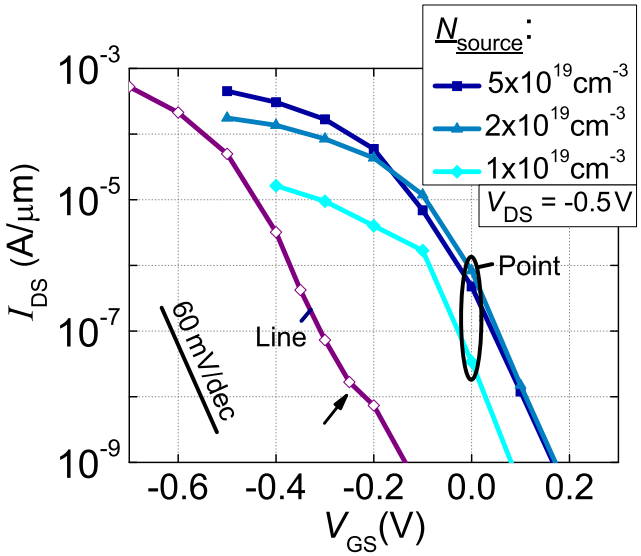

Figure 5.17: QM simulated transfer characteristics of pocketed pointTFETs with optimized T_{po} of 5 nm and varying source doping and a pocketed lineTFET with optimized T_{po} of 5 nm. T_{body} is 20 nm for the pointTFETs and 40 nm for the lineTFET. The arrow indicates a FIQC-induced kink. More details on the configurations can be found in Fig. 5.16 and Table 5.2.

5.6.2 The pTFET issue

Fig. 5.17 shows that neither an optimized pocketed point-pTFET, nor a pocketed line-pTFET reaches sub-60 mV/dec SS, although a 60 mV/dec SS is achieved in both designs. For the lineTFET, the origin of the poor performance lies with the splitting of the heavy hole and light hole subbands in the triangular quantum well underneath the gate-source overlap (see also Section 2.4.2). Tunneling to the heavy hole subbands is less efficient, but starts earlier than tunneling to the light hole subbands. This forms a parasitic tail to the desired transfer characteristics that degrades the SS, which can be seen as a pronounced kink in Fig. 5.17 (indicated with an arrow). If electron-phonon scattering would be included, an additional current tail due to phonon-assisted tunneling to the heavy hole band is expected [42]. For the pointTFET, Fig. 5.18(a) shows that the super-60 mV/dec SS is caused by the large doping degeneracy in the source, which is the result of the low conduction band DOS. The low DOS causes the quasi-Fermi level of the electrons to lie deep inside the band, thereby leaving the exponential tail of the Fermi-Dirac distribution unfiltered. As Fig. 5.17

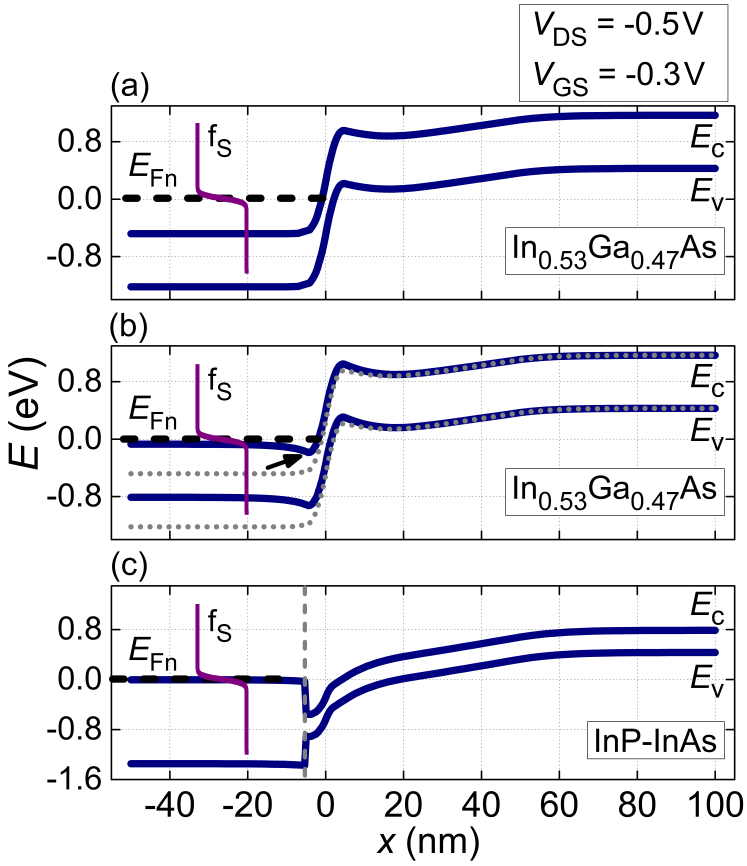

Figure 5.18: Band diagrams along a cutline in the x-direction through the center of the body of three pocketed pointTFET configurations. (a) An n-p-i-p TFET with a T_{po} of 5 nm. (b) An n-n-p-i-p TFET with a T_{po1} and T_{po2} both of 5 nm, with in dotted lines the bands of (a). (c) A heterostructure n-n-p-i-p TFET with a T_{po1} of 5 nm and a T_{po2} of 2 nm and the heterojunction indicated by a vertical dashed line. The origin of the x-axis lies at the junction of the n^+ and p^+-pocket. The Fermi-Dirac distribution f_S and quasi-Fermi level of the electrons (E_{Fn}) are indicated to show the source doping degeneracy. Other configuration details in Fig. 5.16 and Table 5.2.

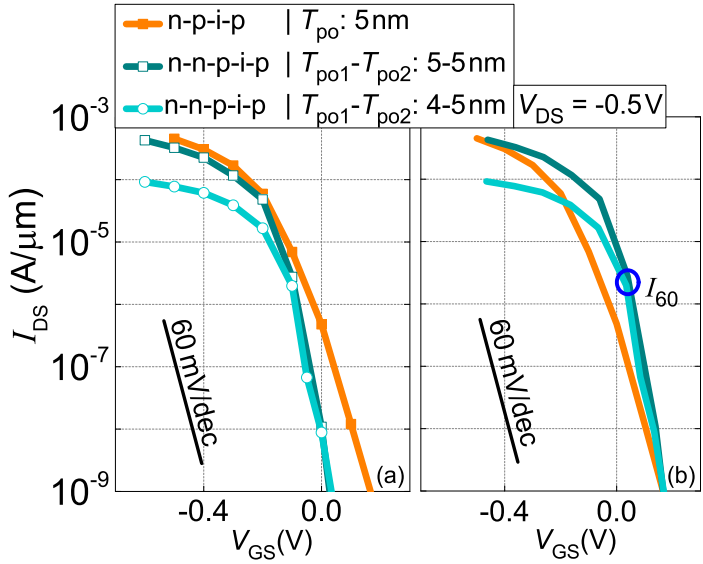

Figure 5.19: QM simulated transfer characteristics of an n-p-i-p pointTFET, compared to two n-n-p-i-p pocketed pointTFETs. T_{body} is 20 nm. (a) Unshifted and (b) shifted characteristics such that V_{GS} at which I_{OFF} is 1 nA/μm coincides in all curves. More configuration details in Fig. 5.16 and Table 5.2.

also shows, simply lowering the source doping is not a solution to the poor performance: in going from 5x10^{19} cm^{-3} to 1x10^{19} cm^{-3} source doping, I_{ON} suffers strongly, as the electric field at the tunnel junction is reduced.

5.6.3 Improved source design

To improve the pTFET SS while retaining I_{ON}, we modify the design of the source. The high doping region is limited to a thin n$^+$-pocket at the tunnel junction, sufficient to locally retain a strong electric field and hence efficient tunneling. In the remainder of the source region, the doping degeneracy is reduced. This can be achieved by reducing the doping level (see Fig. 5.18(b)) or by the introduction of a material with a high conduction band DOS (see Fig. 5.18(c)). The reduced degeneracy in both cases improves the SS, as the tail of the Fermi-Dirac distribution is cut off. We now discuss both options in more detail and address some potential concerns regarding the design.

Applying the concept with a low source doping to the pocketed pointTFET results in an n$^-$-n$^+$-p$^+$-i-n structure as shown in Fig. 5.16(c) with the band

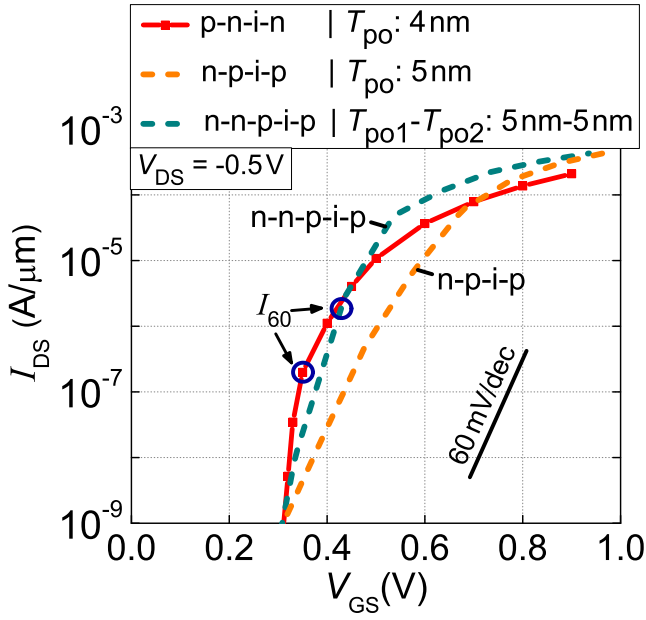

Figure 5.20: QM simulated transfer characteristics comparing an optimized pocketed nTFET with the pTFETs of Fig. 5.19. T_{body} is 20 nm. The pTFET curves are mirrored and shifted such that V_{GS} at which I_{OFF} is 1 nA/μm coincides in all curves. More configuration details in Fig. 5.16 and Table 5.2.

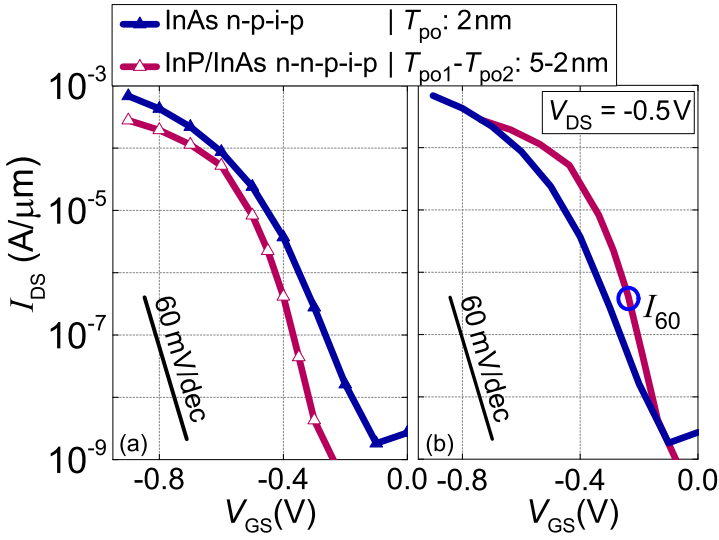

Figure 5.21: QM simulated transfer characteristics of an InAs n-p-i-p pointTFET, compared to an InP-InAs heterostructure n-n-p-i-p TFET. T_{body} is 20 nm. Doping of the InP source region is 5x10^{17} cm^{-3}. (a) Unshifted and (b) shifted characteristics such that V_{GS} at which I_{OFF} is 1 nA/μm coincides in all curves. More configuration details in Fig. 5.16 and Table 5.2.

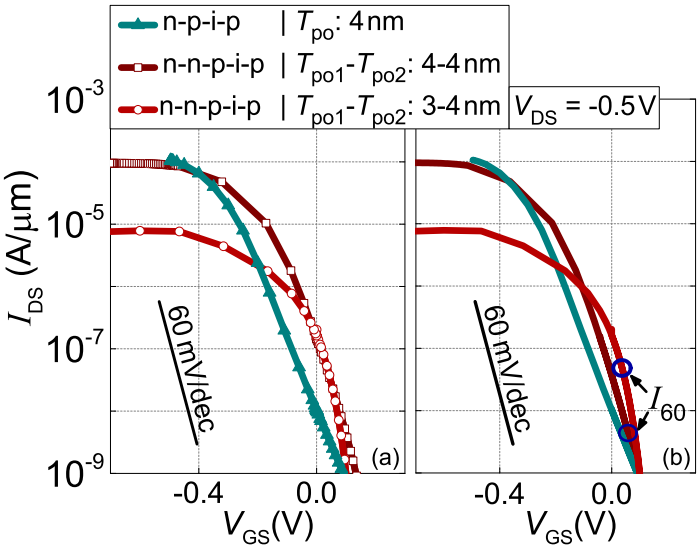

Figure 5.22: SC simulated transfer characteristics of the same configurations as in Fig. 5.19, except for the specifications in the legend. (a) Unshifted and (b) shifted characteristics such that V_{GS} at which I_{OFF} is 1 nA/μm coincides in all curves. More configuration details are found in Fig. 5.16 and Table 5.2.

diagram in Fig. 5.18(b). We will refer to this configuration as the n-n-p-i-p TFET. Fig. 5.19 shows that an n-n-p-i-p configuration with a 5 nm n^+-pocket and a 5 nm p^+-pocket reaches sub-60 mV/dec SS, with an I_{60} of 3 μA/μm, while achieving an I_{ON} comparable to the configuration with a conventional source of 3x10^2 μA/μm for a V_{DD} of 0.5 V and I_{OFF} of 1 nA/μm. In the n-n-p-i-p TFET, T_{po} and gate alignment of both pockets are optimized such that they are almost fully depleted for the full range of V_{GS} (see also Table 5.2). Fig. 5.19 shows, for instance, that for a thinner n^+-pocket, in this case 4 nm, I_{ON} is deteriorated as a result of depletion beyond the pocket into the lowly doped source region at large V_{GS}. This results in a lower electric field at the location of dominant tunneling. Fig. 5.20 shows that a pTFET configuration with an improved source design can match and even surpass an optimized pocketed nTFET with a conventional source design in I_{ON}, depending on the chosen V_{DD}-window. As for the nTFET case, I_{ON} surpasses the target spec of 10^2 μA/μm, while I_{60} is still below the 10 μA/μm target. We will therefore discuss the further optimization of pTFETs in Section 5.7.2.

The improved source design can also be implemented using a heterostructure. The lowly doped source material of Fig. 5.16(d) is now chosen to have a smaller

electron affinity than the remainder of the TFET structure (see Fig. 5.18(c)). The bandgap of the source material cuts off the Fermi-Dirac tail, which again results in a significantly improved SS, while retaining I_{ON} (see Fig. 5.21). Note that to avoid modulation of the energy bands in the source region, a 5 nm gate-source underlap is introduced. Combined with an appropriate choice of InP doping level, this ensures flat energy bands in the source as seen in Fig. 5.18(c).

We now address some potential weaknesses of the improved source design. First, a lowly doped source region, which is technologically the most simple option, may increase the series resistance and access resistance at the source contact. If the latter becomes a problem, however, the high doping can be recovered near the source contact. In the heterostructure case, the material in the source region should ideally have a large conduction band DOS, e.g. like Si, which would allow the predicted improvements to be obtained with a higher source doping and hence a reduced source series resistance. Second, our ballistic QM simulations do not take into account rethermalization, which is caused by interaction of the carriers with lattice phonons. Fast rethermalization can result in a parasitic current tail to our simulated transfer characteristics, as the holes can thermally move to states in the dip of the electrostatic potential at the location of the n^+-pocket (indicated with an arrow in Fig. 5.18(b)). From inside the dip, the holes can then tunnel into the channel. To prevent this, the dip should be kept as small as possible, by choosing the thickness (including any doping gradients) and doping level of the n^+-pocket such that it is nearly depleted. What also helps prevent a rethermalization induced leakage current, is the quantization of the energy levels inside the dip. This reduces the amount of available states for the holes. For instance, in Fig. 5.18(b), the first quantized level is at an energy of 65 meV from the conduction band edge in the source, which is 110 meV from the bottom of the potential dip.

We check with SC simulations that rethermalization indeed does not remove the impact of the improved source design (see Fig. 5.22). The SC simulator assumes the carrier distribution to rethermalize instantly throughout the device to the equilibrium distribution. This also happens in the potential dip, and thus corresponds the worst-case scenario. This condition is opposite to our optimistic ballistic QM simulations, in which the relaxation time of the rethermalization is effectively infinite due to the absence of scattering. The SC simulator also does not consider any quantization effects in the potential dip caused by the n^+- or p^+-pocket. The optimal pocket thickness is therefore smaller, 4 nm in the simulated configurations, to maintain a sufficiently low concentration of free carriers in the n^+-pocket. Similar to the QM case, an n^+-pocket which is too thin results in a decreased I_{ON}. Fig. 5.22 confirms that even in the worst-case scenario of instant rethermalization with no quantization in the potential dip, a significant improvement in SS over the conventional source design is achieved.

5.7 Optimization of heterostructure pointTFETs

While our PhD research focused mainly on solver development and configuration pathfinding (e.g. lineTFET vs pointTFET), our fifteen-band implementation also enabled other members of the imec TFET team to investigate and optimize promising heterostructure TFET configurations [77], with the goal of reaching the performance targets set in Section 2.6. The results of this optimization provided direction to the processing experts and served as input to circuit simulations. Here, we present an example of a few steps in the optimization process of an nTFET (Section 5.7.1) and a pTFET (Section 5.7.2), and we briefly discuss the results of preliminary TFET circuit simulations (Section 5.7.3).

5.7.1 Heterostructure nTFET optimization of T_{body} and T_{po}

Figs. 5.23 and 5.24 show the optimization carried out by A.Verhulst of a staggered bandgap nTFET ($GaAs_{0.5}Sb_{0.5}$/ $In_{0.53}Ga_{0.47}As$, see Fig. 4.4 for the band alignment) with respect to T_{body} and T_{po} [77]. In line with our two-band results in Section 4.4, the simulations show a competing effect between improved gate control and increased SIQC for thinner bodies: compared to the 20 nm case, the SS for the 10 nm T_{body} configurations is improved at the expense of some current at high V_{GS}. For both T_{body} cases, the 3 nm T_{po} configurations perform best overall, with the 10 nm case showing an I_{60} around 1.5x10^{1} μA/μm with an I_{ON} around 1.8x10^{2} μA/μm, while the 20 nm case obtains an I_{60} around 5 μA/μm and an I_{ON} of about 1.2x10^{2} μA/μm for a V_{DD} of 0.3 V and an I_{OFF} of 10^{-11} A/μm. Comparing these results with the targets discussed in Section 2.6 ($I_{60} > 10$ μA/μm and $I_{\mathrm{ON}} > 10^{2}$ μA/μm for a $V_{\mathrm{DD}} < 0.5$ V) show promise for this type of configuration to compete with MOSFET.

5.7.2 Heterostructure pTFET optimization of T_{body} and T_{po}

A similar optimization was carried out for a staggered heterostructure ($In_{0.53}Ga_{0.47}As$/$GaAs_{0.5}Sb_{0.5}$) pTFET configuration, implemented with the improved source design as introduced in Section 5.6.3. In Figs. 5.25 and 5.24, the improved source shows its merit, with values for I_{60} close to, and even surpassing, those of the nTFET, while I_{ON} is above 10^{2} μA/μm for most configurations. The configuration with a T_{po} of 3 nm and a T_{body} of 20 nm shows the best performance with an I_{60} around 2x10^{1} μA/μm and an I_{ON} of 5.3x10^{2} μA/μm with I_{OFF} 10^{-10} A/μm in a 0.3 V V_{DD} window.

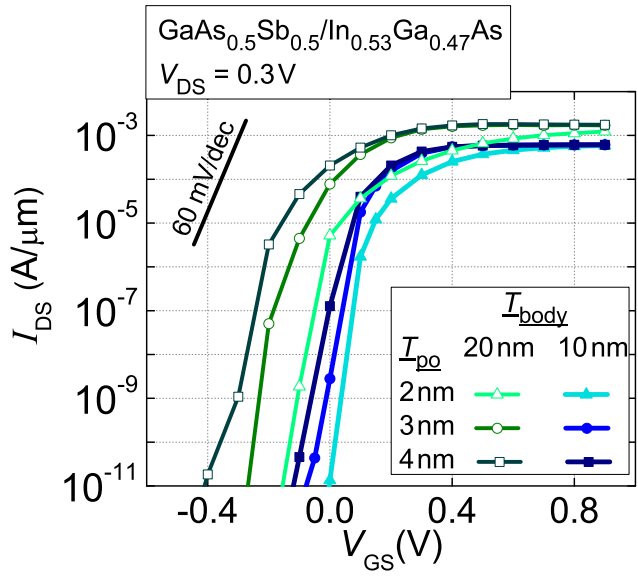

Figure 5.23: QM simulated transfer characteristics of staggered gap heterostructure n-channel pointTFETs with varying T_{body} and T_{po}. The source doping is 1x10^{20} cm^{-3}. Note that the values of I_{DS} above 1 mA/μm are too optimistic because of the absence of series or contact resistance in the simulations. More configuration details in Fig. 5.8(a) and Table 5.1.

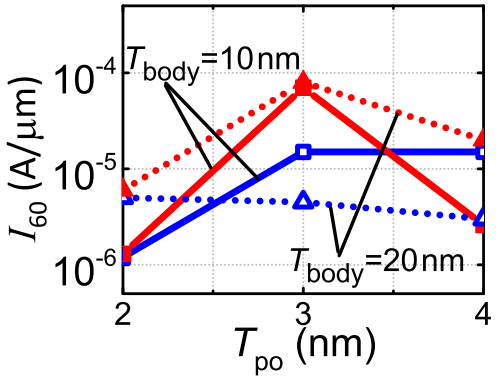

Figure 5.24: QM optimization of staggered gap heterostructure n-channel (open) and p-channel (closed) pointTFETs with varying T_{body} and T_{po}. The source doping is 1x10^{20} cm^{-3}. More configuration details in Figs. 5.8(a) and 5.16(c) and Tables 5.1 and 5.2.

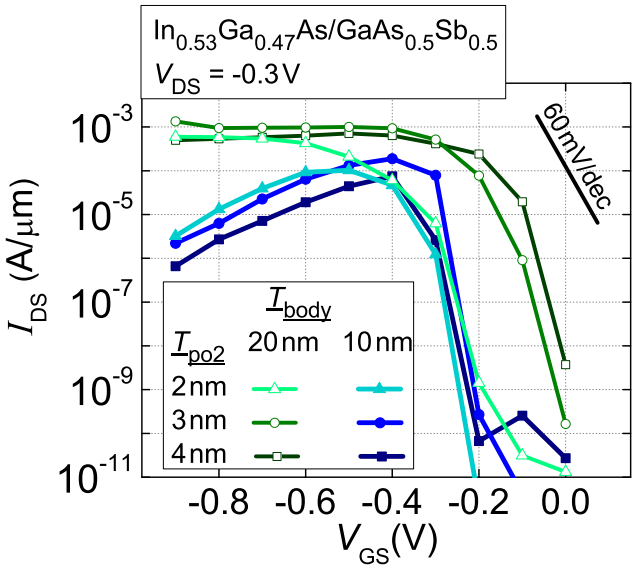

Figure 5.25: QM simulated transfer characteristics of staggered gap heterostructure p-channel pointTFETs with varying T_{body} and T_{po2}. T_{po1} is 2.5 nm in all configurations. Note that the values of I_{DS} above 1 mA/μm are too optimistic because of the absence of series or contact resistance in the simulations. More configuration details in Fig. 5.16(c) and Table 5.2.

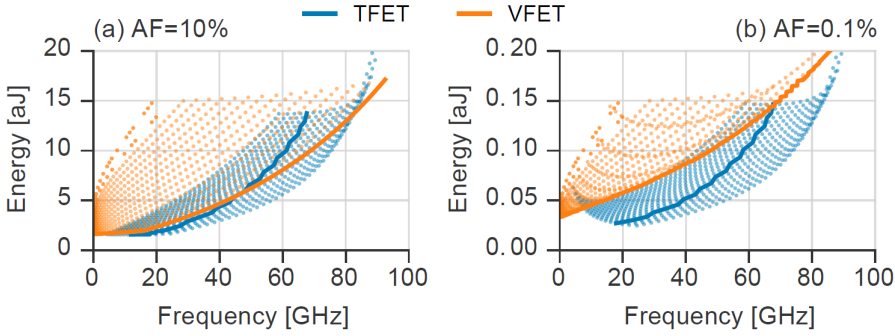

Figure 5.26: Energy versus frequency of a fifteen-stage ring oscillator for two AF values. Each point is a (V_{DD},V_{th})-pair, the solid line is the optimal curve. The ring oscillator stages have a fan-out of 3 and an interconnect length of 50 times the contact gate pitch. V_{DD} range is 0.2 V to 0.6 V. I_{OFF} range is 10 pA to 10 nA. Figure courtesy of D.Yakimets.

5.7.3 Circuit simulations

The optimized transfer characteristics of the previous section, together with output characteristics and capacitances that were also calculated using our QM simulator, have been used as input for preliminary TFET circuit simulations by D.Yakimets [160]. These simulations compare the energy consumption per cycle of a fifteen-stage TFET ring oscillator to the same circuit made up of vertical NW MOSFETs. Fig. 5.26 shows the resulting energy versus frequency curves for two values of the activity factor (AF). For the lower AF of 0.1%, the TFET allows for energy gains up to fifty percent, depending on the oscillator frequency. For the higher AF, the gains are very small. This study indicates that the low SS and leakage current of the TFET allows for energy gains compared to vertical MOSFETs in actual circuits, with the size of the gains dependent on the application.

Note that this is only a preliminary circuit assessment. Although the TFETs were optimized for transfer and output characteristics, this was not yet the case for the device capacitances, which play a large role when switching the device in a circuit.

5.8 Experimental comparison

Besides the modeling and optimization of promising TFET configurations, the imec TFET team has also undertaken an experimental effort to fabricate $In_{0.53}Ga_{0.47}As$ pointTFETs [133]. The structures are not yet the optimized configurations of the previous section, but more simple homostructure test vehicles. In this section, these basic test vehicles allow us to check whether our solver captures the most important device current components. We do this by comparing the experimentally measured transfer characteristics to SC and fifteen-band QM predictions.

5.8.1 Simulation details

The SC and QM simulated configurations are double gate $In_{0.53}Ga_{0.47}As$ p-i-n pointTFETs with a wide T_{body} of 30 nm. Although the experimental devices are single gate, the wide T_{body} ensures there is no coupling between the two gates. We can therefore simply divide the simulated double gate currents by two to obtain the current per gate. The experimental source doping is not exactly known, but from secondary ion mass spectrometry (SIMS) measurements, we estimate it to be around $2x10^{19}$ cm^{-3} [133]. Experimentally, the gate dielectric

consists of 1 nm of Al_2O_3 ($\epsilon = 9$) and 3 nm of HfO_2 ($\epsilon = 25$), resulting in a calculated EOT of 1 nm. V_{DS} is set at 0.2 V. The other configuration details are the same as in Section 5.5, except for the lack of a source pocket, and can be found in Fig. 5.8(a) and Table 5.1. The fifteen-band QM parameters can be found in Appendix C.2.1 and the SC parameters in Appendix D.3.2. The fabrication of the experimental configuration is detailed by Alian *et al.* [133].

5.8.2 Comparison results

Fig. 5.27 shows that the SC simulations overestimate I_{ON} with a full order of magnitude ($4x10^{-2}$ $\mu A/\mu m$ instead of $4x10^{-3}$ $\mu A/\mu m$ in a V_{DD} window of 0.2 V, with an I_{OFF} of $1x10^{-11}$ $A/\mu m$), while the fifteen-band QM simulations match both the SS and I_{ON} of the experimental curve. We attribute the SC overestimation to the lack of FIQC in the SC simulations. FIQC results in a capacitive equivalent thickness (CET) that is significantly larger than the EOT. Based on capacitance-voltage measurements, Alian *et al.* estimated a CET around 1.5-1.7 nm. We find that if the simulated EOT is increased to 2 nm in the SC simulations, they coincide with the QM results.

For higher values of V_{GS}, the QM simulations also start to show a discrepancy with the experiment. This can be due to the contact resistance, which is not included in the simulations.

This comparison shows that the QM simulations give a more realistic prediction of the device performance than the SC simulations, primarily because they capture FIQC. Note that these conclusions are only valid for non-confined, homostructure devices.

5.9 Conclusion

In this chapter, we expanded our two-band formalism to a full-zone fifteen-band implementation. We showed that a straightforward increase in the number of bands is prevented by difficult to remove zone-center spurious solutions and an unrealistic computational burden. We identified the finite difference scheme in the confined direction as the major origin of these issues and therefore replaced it with a spectral approach, in which the EFs are decomposed into a mixed set of sines and cosines. The spectral form of the EF equations allowed us to eliminate zone-center spurious solutions and limit the number of equations, and hence the computational effort.

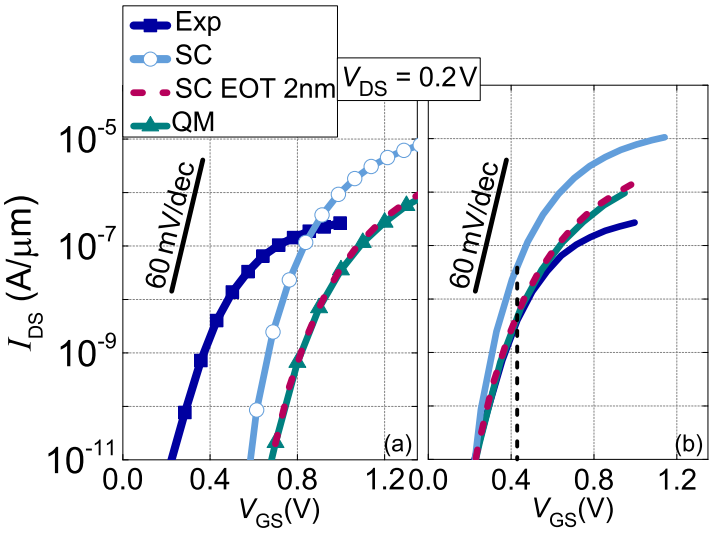

Figure 5.27: Comparison of experimental transfer characteristics to QM and SC simulations. The simulated TFET is a double gate $In_{0.53}Ga_{0.47}As$ homostructure p-i-n pointTFET with a T_{body} of 30 nm. (a) Unshifted and (b) shifted characteristics such that V_{GS} at which I_{OFF} is 10^{-11} A/μm coincides in all curves. The dashed vertical line indicates the V_{DD}-window. The experimental configuration is detailed by Alian *et al.* [133]. More simulated configuration details in Fig. 5.8(a) and Table 5.1.

To enable the simulation of a wide variety of materials, we developed a procedure to obtain fifteen-band material parameters from experimental bandgap energies and effective masses, while conserving the commutativity of the momentum matrices. The commutativity is required for a correct implementation of the heterostructure basis function transformation.

With our fifteen-band implementation and the required parameters, we then confirmed the viability of the lineTFET concept in a direct bandgap III-V material system, but also showed that a properly optimized pocketed pointTFET can obtain similar performance in the case of a relatively small direct bandgap (below 1 eV). Comparing the fifteen-band results with two-band simulations, we found a qualitative match for non-confined pocketed pointTFETs, but large discrepancies for either confined or lineTFET configurations, which therefore require a fifteen-band analysis. For the pTFET, we introduced an improved

source concept that allows for pTFET performance at the same level as nTFET. We illustrated the utility of our simulator for the imec TFET team in optimizing heterostructure nTFET and pTFET for preliminary circuit simulations. Finally, we showed a better match of the fifteen-band simulations to p-i-n homostructure experimental TFETs fabricated at imec than SC simulations.

An important general conclusion from the findings in this chapter is that in a direct bandgap III-V material system with a bandgap below 1 eV, the pointTFET configuration is to be preferred over the lineTFET. An optimized pointTFET can obtain similar performance than the lineTFET, while being easier to fabricate. Additionally, thanks to an improved source design, the pointTFET has the prospect of complementary implementation. The performance of the p-channel lineTFET, on the other hand, is severely degraded by the different quantization of the valence bands in the triangular well under the gate, for which there is no solution at present.

Although the implementation discussed in this chapter is capable of simulating a wide range of configurations, it has two important missing features. One is the lack of spin-orbit coupling, which means the absence of a split-off band. This can become important in situations in which the band structure is warped, e.g. as a consequence of strain. The other is a link between the electrostatic potential and the electron density. This is relevant for TFETs with a very thin body. These two missing features will be added and discussed in the next two chapters.

The spectral approach and simulation results in this chapter have been partly published in the *Journal of Applied Physics* [161]. The improved source concept has been published separately in *Applied Physics Letters* [162].

Chapter 6

Thirty-band model

The simulations of various relaxed homo- and heterojunction TFET configurations in the previous chapters showed promising performance to compete with MOSFET, especially for heterostructure pointTFETs. We now investigate if mechanical strain can further improve the performance of these configurations. As introduced in Section 2.10, strain is an important mobility booster in today's MOSFETs, and is promising for TFET as well, because it allows to modify the semiconductor band structure favorably. Strain can be included in any band model, but because we want to model the intricate effects of strain on the band curvature and DOS, we turn in this chapter to a thirty-band model, which includes spin-orbit band splitting and hence the valence split-off band. In this model, we introduce the effects of strain with a position-dependent zone-center strain Hamiltonian. This allows us to assess both uniform and non-uniform strain profiles in the TFET structure. Uniform strain is the simplest to model and provides insight into which type (uniaxial, biaxial) or sign (tensile, compressive) of stress is beneficial for the TFET performance. However, uniform strain is difficult to obtain experimentally, as it would require an externally applied stress over the full TFET structure. Non-uniform strain, on the other hand, is intrinsically present at a lattice-mismatched heterojunction, but has a more complicated influence on the TFET operation. A proper understanding of the effects of non-uniform strain would open up the TFET design space to lattice-mismatched material combinations and might provide an additional performance tuning parameter.

The chapter is structured as follows. We first describe the thirty-band basis set (Section 6.1). We introduce the basis functions (Section 6.1.1) and compare thirty-band simulations to results of the fifteen-band implementation of the

previous chapter (Section 6.1.2). Next, we discuss the way in which strain is introduced through a strain Hamiltonian (Section 6.1.3). We then apply the thirty-band implementation to uniformly strained lattice-matched structures, and show how the source concept introduced in Section 5.6.3 enables us to take full benefit of the strain-induced I_{ON} gain (Section 6.2). Next, we assess whether a non-uniform strain profile at a lattice-mismatched heterojunction can improve TFET performance and provide an alternative to dopant pockets or high source doping (Section 6.3). The chapter is summarized in the conclusion section (Section 6.4).

6.1 Thirty-band basis set

6.1.1 Choice of basis functions

The thirty-band model is essentially the fifteen-band model of Chapter 5 with the inclusion of electron spin, which is a form of intrinsic angular momentum. The amount of basis states therefore doubles, since every state now has a spin up and spin down version, e.g. for the first conduction and valence band states we have:

$$\begin{aligned} &|S\uparrow\rangle\,|S\downarrow\rangle \\ &|X\uparrow\rangle\,|X\downarrow\rangle\,|Y\uparrow\rangle\,|Y\downarrow\rangle\,|Z\uparrow\rangle\,|Z\downarrow\rangle \end{aligned} \tag{6.1}$$

and so on for the other states. The basis set is schematically summarized in Fig. 6.1.

With the inclusion of spin, spin-orbit coupling splits some of the energy levels that are degenerate in a fifteen-band model. Spin-orbit coupling is a relativistic interaction between the spin of the electron and the magnetic field that the electron experiences in its rest frame in orbit around the nucleus [163]. The interaction separates states in energy depending on the alignment of their spin vector with the orbital magnetic field. The most notable example is the split-off valence band, which is formed by one of the two degenerate heavy hole bands of the fifteen-band model splitting off in energy. Similar spin splittings exist between the other p-like and d-like states. These splitting energies are indicated in Fig. 6.1 as Δ, Δ_{c} and Δ_{d}. Expressions for these energies can be determined starting from the spin-orbit $\mathbf{k}\cdot\mathbf{p}$-term:

$$\frac{\hbar}{4c^2 m_e^2}(\nabla V_c \times \mathbf{p}) \cdot \boldsymbol{\sigma} \tag{6.2}$$

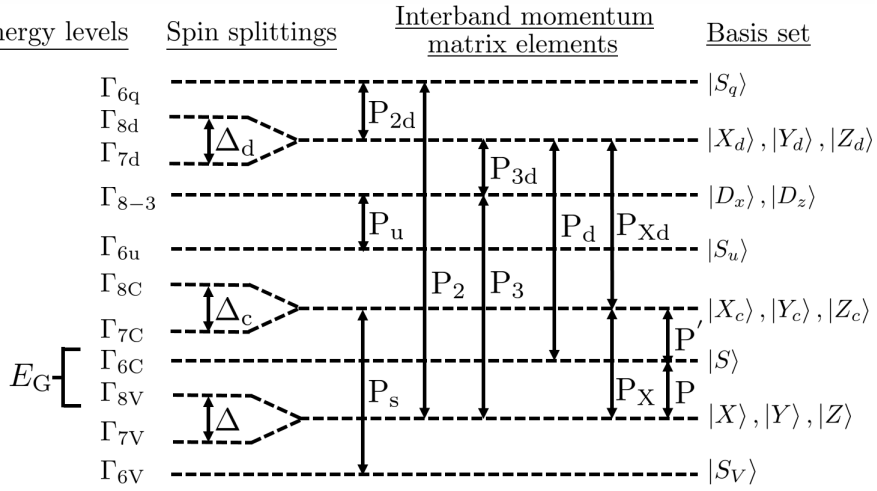

Figure 6.1: Energy levels, spin-orbit splittings and interband momentum matrix elements of the chosen thirty-band basis set [58]. Each basis state has a spin up and spin down version.

with c the speed of light in vacuum, V_c the crystal potential, $\mathbf{p}$ the momentum operator and the components of $\boldsymbol{\sigma}$ the Pauli spin matrices:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \tag{6.3}$$

The spin-orbit energy splittings are then the matrix elements of Eq. (6.2) with the directional basis states (p-like and d-like). Using the symmetry properties of the basis states, Δ, Δ_c and Δ_d can be written as [58]:

$$\begin{aligned} \Delta &= \frac{3\hbar}{4m_e^2c^2} \langle X| \frac{\partial V_\mathrm{c}}{\partial x} p_y - \frac{\partial V_\mathrm{c}}{\partial y} p_x |iY\rangle \\ \Delta_\mathrm{c} &= \frac{3\hbar}{4m_e^2c^2} \langle X_\mathrm{c}| \frac{\partial V_\mathrm{c}}{\partial x} p_y - \frac{\partial V_\mathrm{c}}{\partial y} p_x |iY_\mathrm{c}\rangle \\ \Delta_\mathrm{d} &= \frac{3\hbar}{4m_e^2c^2} \langle X_\mathrm{d}| \frac{\partial V_\mathrm{c}}{\partial x} p_y - \frac{\partial V_\mathrm{c}}{\partial y} p_x |iY_\mathrm{d}\rangle \end{aligned} \tag{6.4}$$

For transport simulations, Δ is the most important of these three energies, as it is the splitting energy between the split-off band and the heavy and light hole bands, which determine the transport of holes. Δ can be determined from optical measurements [164], while the other spin splitting energies are either

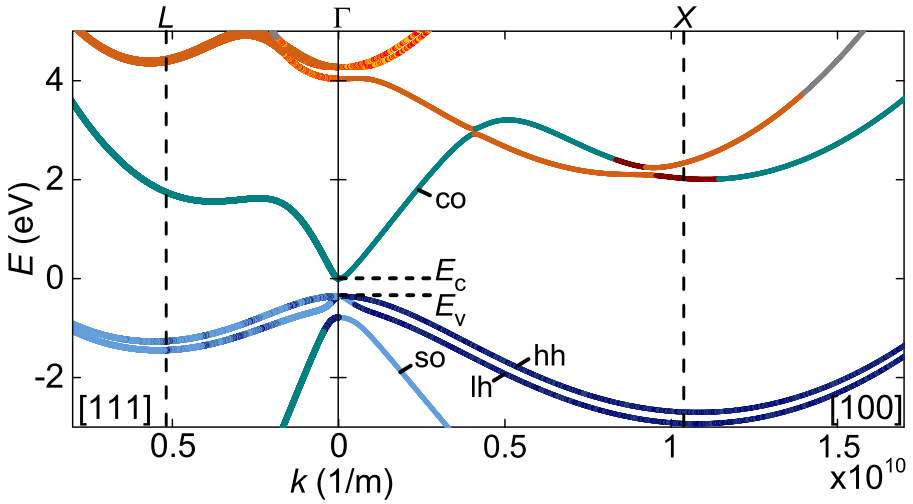

Figure 6.2: Thirty-band bulk band structure of InAs in the [100] transport direction (right) and the [111]-direction (left). The dashed lines indicate the edge of the first Brillouin zone. The solutions are colored based on the dominant basis state. Indicated are the conduction (co), light hole (lh), heavy hole (hh) and split-off (so) bands. The thirty-band parameters are taken from Radhia *et al.* [58] and are listed in Appendix C.2.

obtained from ab-initio calculations [165], or act as additional band structure fitting parameters. In our implementation, we use Δ_c as a fitting parameter in the parameter search algorithm of Section 5.4 and set Δ_d to zero, as it only influences bands that are very high in energy. Theoretically, several other spin splitting energies exist between the different directional p-like and d-like states, but these are so small that we can safely neglect them for our purposes [58, 150].

The spin splittings of Eq. (6.4) are added as an energy term to the EF system. They are collected in a spin-orbit matrix $\mathbf{H}_{\mathrm{SO}}$, which is then included in the term describing the bulk interband Hamiltonian matrix elements in the spectral EF system of Eq. (5.8):

$$\sum_m \left[H^k_{nm}(x) + [H^k_{\mathrm{SO}}]_{nm}(x)\right] \tilde{F}^k_m(x, k_{z\mu}) = \sum_m [H^k_{\mathrm{cmb}}]_{nm}(x) \tilde{F}^k_m(x, k_{z\mu}) \quad (6.5)$$

with $[H^k_{\mathrm{SO}}]_{nm}$ the spin-orbit matrix elements, which have been transformed to the set of basis functions of material layer k according to Eq. (3.9), and $[H^k_{\mathrm{cmb}}]_{nm}$ the resulting combined matrix element. The construction of the spin-orbit matrix $\mathbf{H}_{\mathrm{SO}}$ is defined in Appendix A.5. We add a completely analogous term to the contact eigenvalue problem of Eq. (5.9).

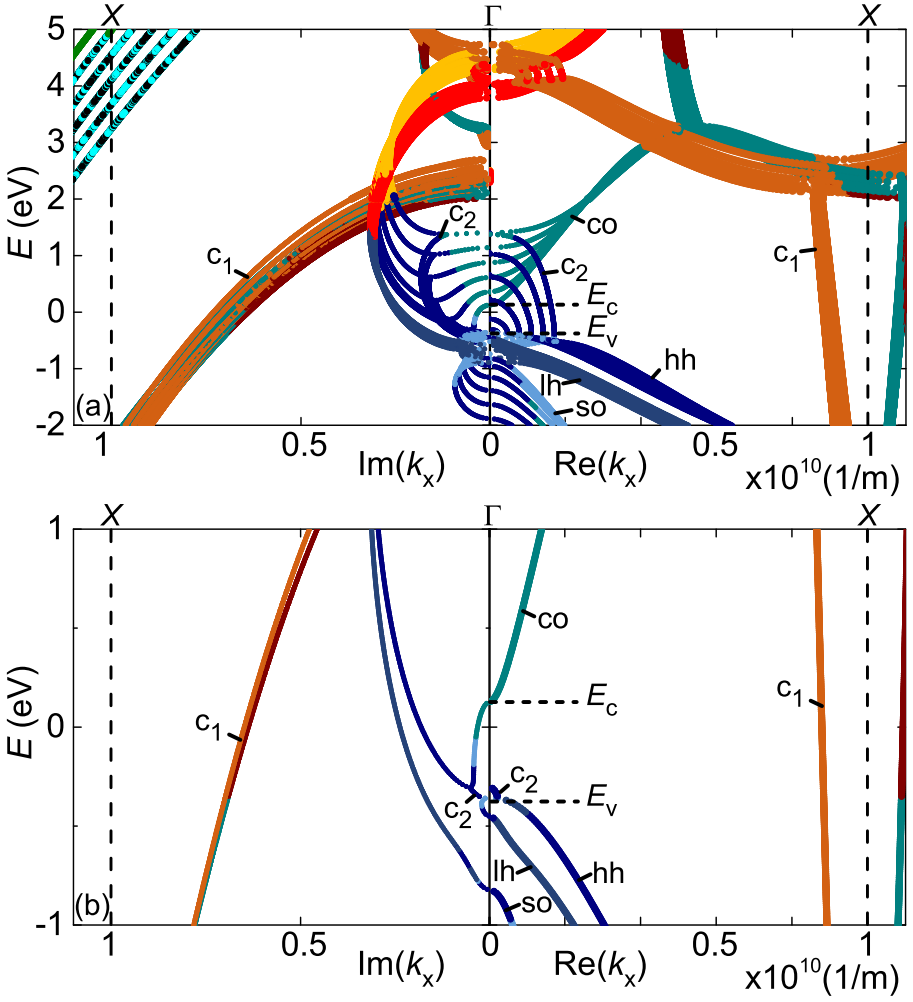

Figure 6.3: Thirty-band confined real (right) and imaginary (left) band structure of a 10 nm InAs slab in the transport direction, calculated using a spectral method in the confined direction with (a) 6 and (b) 1 spectral components. k_y is zero. The dashed lines indicate the edge of the first Brillouin zone. The solutions are colored based on the dominant basis state. Indicated are the conduction (co), light hole (lh), heavy hole (hh) and split-off (so) bands, and two types of complex solutions (c_1 and c_2). The thirty-band parameters are taken from Radhia *et al.* [58] and are listed in Appendix C.2.5.

Fig. 6.2 shows an example of a bulk thirty-band band structure of InAs. All bands are now explicitly spin degenerate. Comparing to the fifteen-band case in Fig. 5.2, the most important changes occur in the valence band: a split-off band appears, while the strong curvature of the light hole band is limited to a small region around the zone center of about 0.3 eV. The change in light hole band curvature from strong to weak has an impact both on the valence band DOS and the quantization of the subbands when confined (see Fig. 6.3 and compare with Fig. 5.5).

6.1.2 Comparison fifteen-band to thirty-band simulations

Just as we did for the two-band and fifteen-band models in Section 5.5.3, we now compare the fifteen-band to the thirty-band model with current simulations to see the effect of the differences in band structure. We revisit the comparison for $In_{0.53}Ga_{0.47}As$ p-i-n diodes and pocketed pointTFETs with varying degree of confinement. The thirty-band parameters are obtained with the fitting procedure discussed in Section 5.4 and are listed in Appendix C.2.1.

Fig. 6.4 shows for the same $In_{0.53}Ga_{0.47}As$ diodes as in Fig. 5.11 that the thirty-band and fifteen-band results match very well down to a T_{body} of 10 nm. For a T_{body} of 5 nm, the strong confinement results in a larger discrepancy. As in the two-band to fifteen-band comparison, the explanation lies in the differences in band structure. The discrepancy is largest for small values of V_{DS}, when the tunneling window is small. At this bias point, the difference in DOS at the top of the valence band plays a significant role.

Fig. 6.5 shows the comparison for the different band models for the pocketed pointTFETs of Fig. 5.12. Similar to the diode results, the three models give a qualitatively similar result for a T_{body} of 10 nm, but differ strongly for the 5 nm configurations as a result of differences in the band structure. This indicates that for quantitative simulations of very confined structures, the calibration of the band structure parameters should go beyond the zone center or select points of symmetry. This would require the optimization of the parameters compared to an ab-initio reference and is beyond the scope of this work. In the remainder of the chapter, we therefore continue the approach of the previous chapter to only consider configurations with a T_{body} of 10 nm or larger.

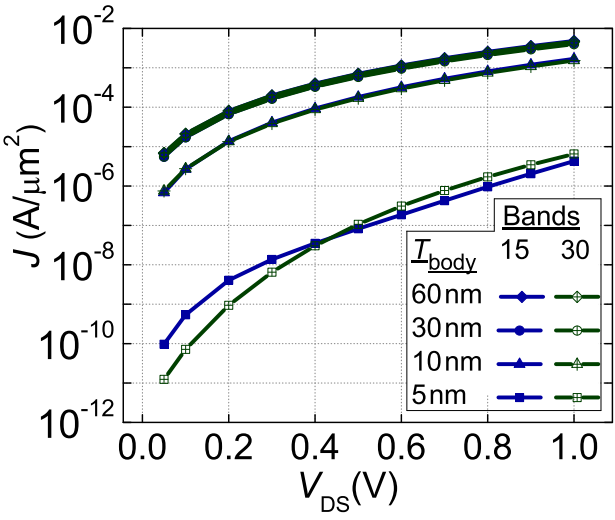

Figure 6.4: Fifteen- and thirty-band QM simulations of current densities for the $In_{0.53}Ga_{0.47}As$ p-i-n diode of Fig. 4.3 with a T_i of 9 nm and varying T_{body}. The electrostatic potential for a given T_{body} is identical for both band models.

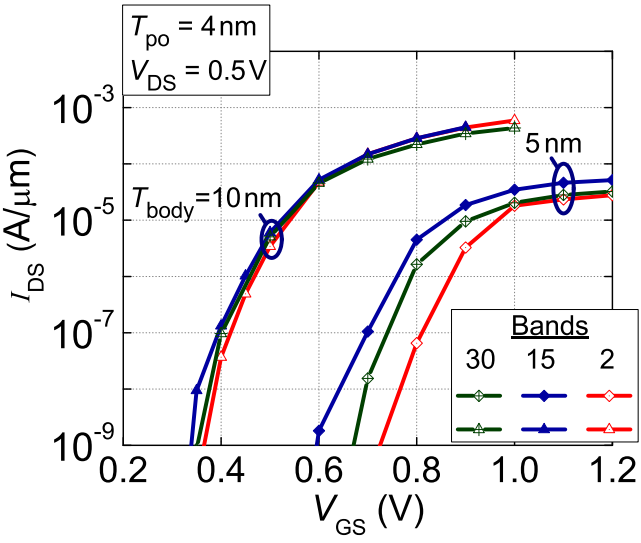

Figure 6.5: Two-, fifteen- and thirty-band QM simulated transfer characteristics of a pocketed pointTFET with two values for T_{body}. The electrostatic potential for a given T_{body} is identical for all band models. More details on the configurations can be found in Fig. 5.8(a) and Table 5.1.

6.1.3 Strain Hamiltonian

Since the aim of this chapter is to investigate the effect of strain on the TFET performance, we complement the thirty-band model of the previous section with a strain Hamiltonian. The construction of this strain Hamiltonian starts with the infinitesimal strain tensor for a general non-uniform strain profile:

$$\overline{\overline{\varepsilon}}(\mathbf{r}) = \begin{bmatrix} \varepsilon_{xx}(\mathbf{r}) & \varepsilon_{xy}(\mathbf{r}) & \varepsilon_{xz}(\mathbf{r}) \\ \varepsilon_{yx}(\mathbf{r}) & \varepsilon_{yy}(\mathbf{r}) & \varepsilon_{yz}(\mathbf{r}) \\ \varepsilon_{zx}(\mathbf{r}) & \varepsilon_{zy}(\mathbf{r}) & \varepsilon_{zz}(\mathbf{r}) \end{bmatrix} \tag{6.6}$$

which contains both normal (ε_{xx}, ε_{yy}, ε_{zz}) and shear (ε_{xy}, ε_{yx}, ε_{yz}, ε_{zy}, ε_{xz}, ε_{zx}) strain components. The elements of the strain tensor are linked to the mechanical stress components at each point in the material through Hooke's Law for zinc-blende materials [166]:

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \\ \sigma_{xy} \\ \sigma_{xz} \\ \sigma_{yz} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{bmatrix} \begin{bmatrix} \varepsilon_{xx} \\ \varepsilon_{yy} \\ \varepsilon_{zz} \\ 2\varepsilon_{xy} \\ 2\varepsilon_{xz} \\ 2\varepsilon_{yz} \end{bmatrix} \tag{6.7}$$

with C_{11}, C_{12} and C_{44} the elements of the stiffness matrix, available in literature for various III-V materials [81]. Based on the strain tensor of Eq. (6.6), we construct a strain Hamiltonian $\mathbf{H}_{\mathrm{S}}$ as proposed by Bahder [167] for an eight-band model and an isotropic Γ-valley:

$$\mathbf{H}_{\mathrm{S}}(\mathbf{r}) = \begin{bmatrix} \mathbf{A}(\mathbf{r}) & \mathbf{0} \\ \mathbf{0} & \mathbf{A}(\mathbf{r}) \end{bmatrix}$$

$$\mathbf{A}(\mathbf{r}) = \begin{bmatrix} a_{\mathrm{c}}\mathrm{Tr}(\overline{\overline{\varepsilon}}) & 0 & 0 & 0 \\ 0 & f(\overline{\overline{\varepsilon}}) + g(r_{\mathrm{b}}) & (3^{1/2}(r_{\mathrm{b}} - r_{\mathrm{b}}^{*})i)/2 & -(3^{1/2}(s + s^{*}))/2 \\ 0 & (3^{1/2}(r_{\mathrm{b}} - r_{\mathrm{b}}^{*})i)/2 & f(\overline{\overline{\varepsilon}}) - g(r_{\mathrm{b}}) & -(3^{1/2}(s - s^{*})i)/2 \\ 0 & -(3^{1/2}(s + s^{*}))/2 & -(3^{1/2}(s - s^{*})i)/2 & a_{\mathrm{v}}\mathrm{Tr}(\overline{\overline{\varepsilon}}) - 2b_{\mathrm{v}}\varepsilon_{\perp\parallel} \end{bmatrix} \tag{6.8}$$

with

$$
\begin{aligned}
f(\bar{\bar{\varepsilon}}) &= a_{\mathrm{v}}\mathrm{Tr}(\bar{\bar{\varepsilon}}) + b_{\mathrm{v}}\varepsilon_{\parallel\perp} \\
\varepsilon_{\parallel\perp} &= \frac{1}{2}(\varepsilon_{yy} + \varepsilon_{zz}) - \varepsilon_{xx} \\
g(r_{\mathrm{b}}) &= (3^{1/2}r_{\mathrm{b}})/2 + (3^{1/2}r_{\mathrm{b}}^{*})/2 \\
r_{\mathrm{b}} &= \frac{\sqrt{3}}{2}b_{\mathrm{v}}(\varepsilon_{zz} - \varepsilon_{yy}) - id_{\mathrm{v}}\varepsilon_{yz} \\
s &= -d_{\mathrm{v}}(\varepsilon_{xz} - i\varepsilon_{xy})
\end{aligned}
\tag{6.9}
$$

with Tr the trace operator and where we have added a subscript to r_{b} to avoid confusion with the position variable $\mathbf{r}$, while staying consistent with the notation of Bahder. a_{c} and a_{v} are the hydrostatic deformation potential constants of conduction and valence band respectively, while b_{v} and d_{v} are shear deformation potential constants of the valence band. Values for a_{c}, a_{v}, b_{v} and d_{v} are listed in literature for most III-V materials [139]. In our sign convention a_{c} and b_{v} are positive and a_{v} and d_{v} are negative, such that a positive (negative) strain is compressive (tensile). We have dropped the position dependence for the matrix elements in Eqs. (6.8) and (6.9) for clarity of notation, but note that for a general strain in a heterostructure, both the strain tensor components and the deformation potentials have a position dependence.

The use of $\mathbf{H}_{\mathrm{S}}$ of Eq. (6.8) entails a number of approximations. First, because $\mathbf{H}_{\mathrm{S}}$ is only an eight-band Hamiltonian, strain only affects the states in the thirty-band model which are also present in the eight-band model, meaning the first conduction and valence band states. This is an acceptable approximation, since the main effect of strain on the DOS and the tunneling is captured by the bands associated to these states (conduction, light hole, heavy hole and split-off band). Second, like Neffati *et al.*, we have neglected any linear k-dependent terms in Eq. (6.8) as they would significantly complicate the resulting EF system [168]. $\mathbf{H}_{\mathrm{S}}$ is therefore only valid close to the zone center.

The matrix elements of the strain Hamiltonian are now added to the bulk Hamiltonian term of the 2D EF system of Eq. (3.15):

$$
\begin{aligned}
\sum_{m} \left[H^{k}_{nm}(x) + [H^{k}_{\mathrm{SO}}]_{nm}(x) + [H^{k}_{\mathrm{S}}]_{nm}(x,z)\right] F^{k}_{m}(x,z) \\
= \sum_{m} [H^{k}_{\mathrm{cmb}}]_{nm}(x,z) F^{k}_{m}(x,z)
\end{aligned}
\tag{6.10}
$$

which introduces a position dependence in the confined z-direction, on top of the existing position dependence in x because of a possible heterostructure. The position dependence in z means that to follow the spectral approach of Section 5.3, we have to expand the bulk Hamiltonian term on the spectral basis set, like we did for the potential energy term in Eq. (5.5). Analogously to the potential energy, we choose a cosine set to allow for any value of the bulk Hamiltonian term at the confined boundaries:

$$[H^k_{\mathrm{cmb}}]_{nm}(x,z) = \sum_{\mu=-\infty}^{+\infty} [\tilde{H}^k_{\mathrm{cmb}}]_{nm}(x,k_{z\mu})\cos(k_{z\mu}z) \tag{6.11}$$

Through a completely analogous derivation as the potential energy term (see Appendix B), we obtain the spectral form of the bulk Hamiltonian term and include it in the spectral EF system of Eq. (5.8):

$$\begin{aligned}
&\frac{-\hbar^2}{2m_e}\frac{\partial^2 \tilde{F}^k_n(x,k_{z\mu})}{\partial x^2} + k_y^2\frac{\hbar^2}{2m_e}\tilde{F}^k_n(x,k_{z\mu}) + k_{z\mu}^2\frac{\hbar^2}{2m_e}\tilde{F}^k_n(x,k_{z\mu}) \\
&- \frac{i\hbar}{m_e}\sum_m p^{k,x}_{nm}(x)\frac{\partial \tilde{F}^k_m(x,k_{z\mu})}{\partial x} + k_y\frac{\hbar}{m_e}\sum_m p^{k,y}_{nm}(x)\tilde{F}^k_m(x,k_{z\mu}) \\
&\pm k_{z\mu}\frac{i\hbar}{m_e}\sum_m p^{k,z}_{nm}(x)\tilde{F}^k_m(x,k_{z\mu}) \\
&+ \sum_{\mu'=0}^{+\infty}\sum_m \left[[\tilde{H}^k_{\mathrm{cmb}}]_{nm}(x,k_{z\mu'}-k_{z\mu}) \mp [\tilde{H}^k_{\mathrm{cmb}}]_{nm}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}^k_m(x,k_{z\mu'}) \\
&+ \sum_{\mu'=0}^{+\infty}\left[\tilde{V}_e(x,k_{z\mu'}-k_{z\mu}) \mp \tilde{V}_e(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}^k_n(x,k_{z\mu'}) = E\tilde{F}^k_n(x,k_{z\mu})
\end{aligned} \tag{6.12}$$

which can be implemented in the same fashion as Eq. (5.8). The construction of the discretized form of this system is detailed in Appendix A.3. A similar term is included in the spectral contact eigenvalue problem of Eq. (5.9):

$$\sum_{\mu'=0}^{+\infty}\sum_m \left[[\tilde{H}_{\mathrm{cmb}}]_{nm}(x_{\mathrm{c}},k_{z\mu'}-k_{z\mu}) \mp [\tilde{H}_{\mathrm{cmb}}]_{nm}(x_{\mathrm{c}},k_{z\mu'}+k_{z\mu})\right]\tilde{\chi}_{m\alpha}(k_{z\mu'}) \tag{6.13}$$

with x_{c} the location of the contact as before.

In Section 6.2 of this chapter, we discuss the important special case of uniform normal stress. The described general strain tensor and corresponding

strain Hamiltonian can then be simplified. The strain tensor has no position dependence or shear components and of the three remaining components, two are equal. Eq. (6.6) then reduces to:

$$\overline{\overline{\varepsilon}} = \begin{bmatrix} \varepsilon_{\perp} & 0 & 0 \\ 0 & \varepsilon_{\parallel} & 0 \\ 0 & 0 & \varepsilon_{\parallel} \end{bmatrix} \tag{6.14}$$

with $\varepsilon_{\perp}$ and $\varepsilon_{\parallel}$ the strain components in respectively the x and y,z-directions. Eq. (6.7) is therefore simplified to:

$$\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{zz} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{12} \\ C_{12} & C_{11} & C_{12} \\ C_{12} & C_{12} & C_{11} \end{bmatrix} \begin{bmatrix} \varepsilon_{\perp} \\ \varepsilon_{\parallel} \\ \varepsilon_{\parallel} \end{bmatrix} \tag{6.15}$$

In Eq. (6.8) for the strain Hamiltonian, r_{b} and s are zero, resulting in a diagonal matrix:

$$\mathbf{A} = \begin{bmatrix} a_{\mathrm{c}}\mathrm{Tr}(\overline{\overline{\varepsilon}}) & 0 & 0 & 0 \\ 0 & f(\overline{\overline{\varepsilon}}) & 0 & 0 \\ 0 & 0 & f(\overline{\overline{\varepsilon}}) & 0 \\ 0 & 0 & 0 & a_{\mathrm{v}}\mathrm{Tr}(\overline{\overline{\varepsilon}}) - 2b_{\mathrm{v}}(\varepsilon_{\parallel} - \varepsilon_{\perp}) \end{bmatrix} \tag{6.16}$$

The resulting Hamiltonian matrix elements in the EF system therefore have no z-dependence, simplifying the corresponding term in the spectral EF system of Eq. (6.12) to:

$$\sum_{m} [H^{k}_{\mathrm{cmb}}]_{nm}(x) \tilde{F}^{k}_{m}(x, k_{z\mu}) \tag{6.17}$$

6.2 Uniform strain in lattice-matched heterostructure TFETs

With the implemented thirty-band model and strain Hamiltonian, we now assess whether uniform strain can improve the transfer characteristics of lattice-matched heterostructure TFETs, such as the ones in Section 5.7.1. In literature, the effect of strain on the tunneling current has been simulated [108, 169] and measured [15, 16, 170, 171] in group IV configurations, with biaxial tensile stress, in the two directions orthogonal to transport, turning out beneficial for I_{ON} and SS. For III-V TFETs, research based on eight-band $\mathbf{k}\cdot\mathbf{p}$ models has predicted a beneficial effect on I_{ON} for biaxial tensile stress, both in InAs homostructure [17] and broken-gap GaSb/InAs heterostructure TFETs [172].

However, in both cases the SS was found to deteriorate with increasing tensile stress. In this section, we verify these findings, explain the origin of the SS degradation and suggest a way to counteract it.

6.2.1 Simulation details

The simulated configurations are lattice-matched staggered bandgap $GaAs_{0.5}Sb_{0.5}$ /$In_{0.53}Ga_{0.47}As$ pocketed pointTFETs, such as the ones optimized in Section 5.7.1. The configurations are displayed in Fig. 6.6 with details in Table 6.1. The thirty-band QM parameters can be found in Appendices C.2.1 ($In_{0.53}Ga_{0.47}As$) and C.2.2 ($GaAs_{0.5}Sb_{0.5}$), and the SC material parameters used for the calculation in SDevice of the electrostatic potentials in Appendices D.3.2 ($In_{0.53}Ga_{0.47}As$) and D.3.3 ($GaAs_{0.5}Sb_{0.5}$).

To these configurations, we apply 500 MPa of uniform tensile and compressive uniaxial stress, and a tensile biaxial stress. For the uniaxial case, $\sigma_{xx} = \pm 500$ MPa and $\sigma_{yy} = \sigma_{zz} = 0$ and for the biaxial case $\sigma_{xx} = 0$ and $\sigma_{yy} = \sigma_{zz} = 500$ MPa. These values are in line with experimentally achievable values of uniform stress. The uniaxial stress could for example be realized in suspended nanowires [173, 174], while the biaxial tensile stress could be the result of lattice mismatch with an underlying substrate. Based on the stress values, we calculate the corresponding strain tensor elements with Eq. (6.15). Since the strain tensor is uniform as in Eq. (6.14), we use the simplified $\mathbf{H}_S$ of Eq. (6.16) to include the strain in the EF system.

We obtain the electrostatic potential from a SC calculation with SDevice as discussed in Section 4.4.1. To incorporate the effect of the uniform strain on the Fermi level and hence the potential, we first calculate the curvature effective masses at the Γ-point with a bulk strained 30-band $\mathbf{k}\cdot\mathbf{p}$ model, and insert those into the SC simulator.

6.2.2 Results

Fig. 6.7 shows that, in agreement with literature, the main effect of the uniform strain is a deterioration in SS, with the biaxial case reaching higher currents for high V_{GS} than the other strain conditions. The band diagrams of Fig. 6.8(a) show that the SS degradation is caused by an increase in source degeneracy for all strained configurations. As discussed in Section 2.3, a larger source degeneracy leaves the exponential tail of the Fermi-Dirac distribution partially uncovered, which degrades the SS.

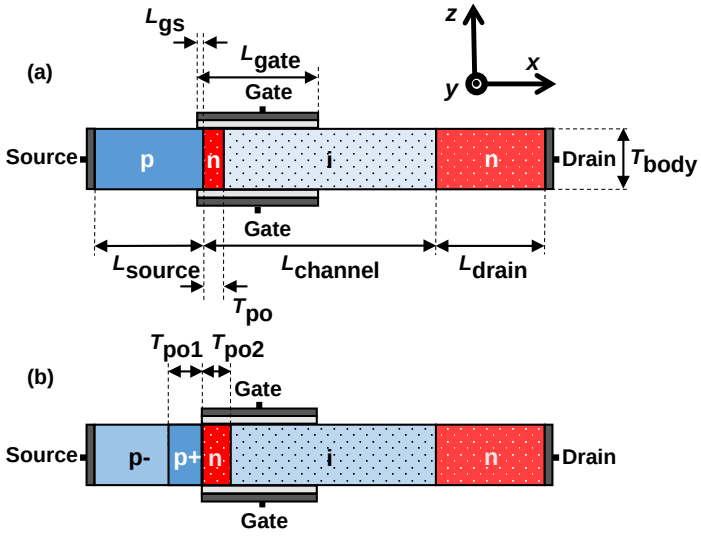

Figure 6.6: Investigated nTFET configurations: (a) pocketed heterostructure pointTFET and (b) pocketed heterostructure pointTFET with improved source design. The solid and dotted regions consist of respectively $GaAs_{0.5}Sb_{0.5}$ and $In_{0.53}Ga_{0.47}As$.

Table 6.1: Details of simulated nTFET configurations (see Fig. 6.6).

L_{source} [nm]	20
L_{channel} [nm]	50
L_{drain} [nm]	15
L_{gate} [nm]	21
L_{gs} [nm]	1
T_{body} [nm]	10
T_{po} [nm]	varying
Doping p-source [cm^{-3}]	$5\text{x}10^{19}$
Doping p^{-}-source [cm^{-3}]	$5\text{x}10^{18}$
Doping n-pocket [cm^{-3}]	$5\text{x}10^{19}$
Doping p^{+}-pocket [cm^{-3}]	$5\text{x}10^{19}$
Doping channel [cm^{-3}]	$1\text{x}10^{10}$
Doping drain [cm^{-3}]	$5\text{x}10^{17}$
EOT [nm]	0.6
GWF [eV]	5

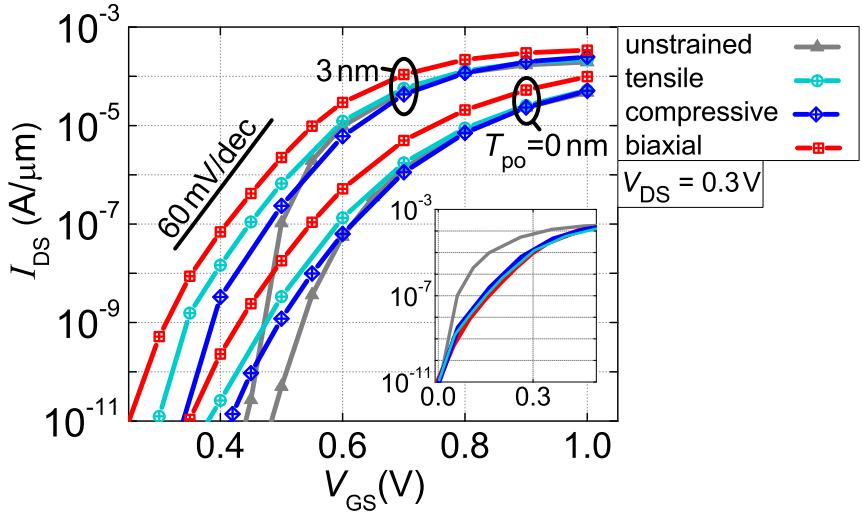

Figure 6.7: QM transfer characteristics of uniformly stressed heterostructure pocketed pointTFETs for a T_{po} of 3 nm and 0 nm. Inset: 3 nm curves shifted to the same I_{OFF}. More configuration details in Fig. 6.6 and Table 6.1.

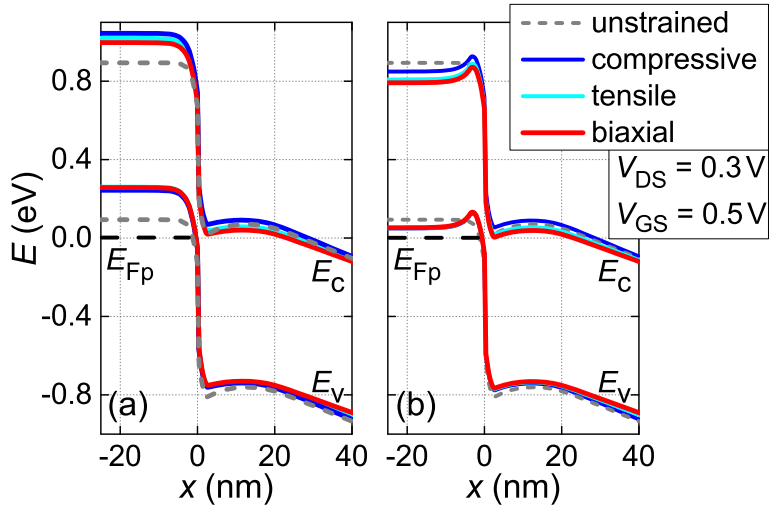

Figure 6.8: Confined band diagrams at the tunnel junction for different strain types along a body center cutline through (a) a p-n-i-n TFET with a T_{po} of 3 nm and (b) a p-p-n-i-n TFET with a T_{po1} of 4 nm and a T_{po2} of 3 nm. The dashed black horizontal line shows the hole Fermi level (E_{Fp}). The grey dashed lines represent the band diagram of the reference unstrained p-n-i-n TFET.

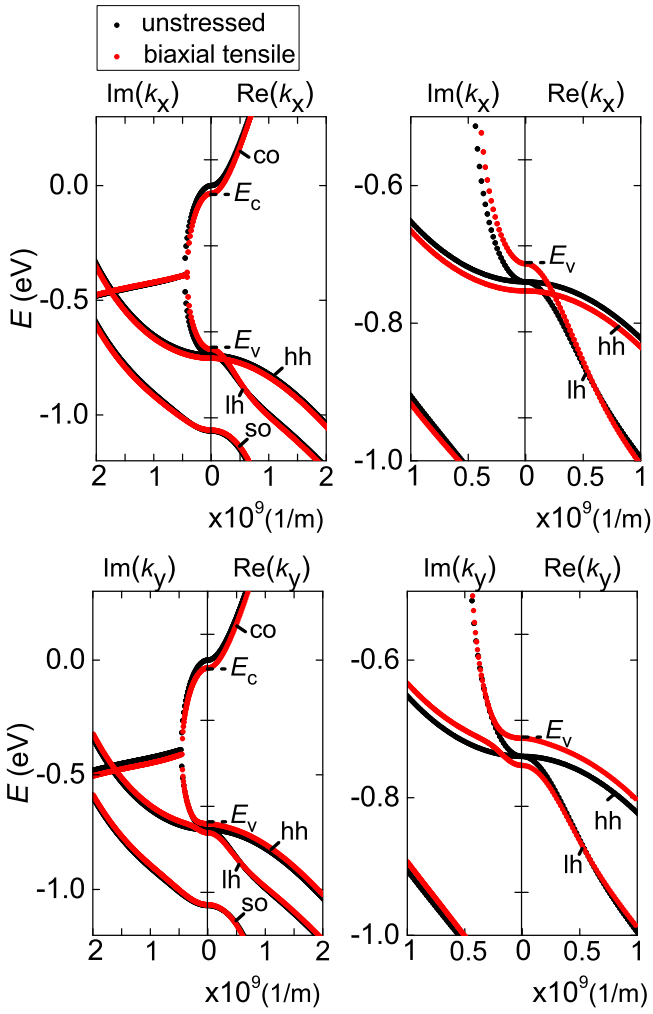

Figure 6.9: Unstrained and strained real and imaginary bulk band structure of $In_{0.53}Ga_{0.47}As$ in the [100] transport direction and the [010] orthogonal direction. The right panels are a zoom of the valence band edge. k_z is zero. Indicated are the conduction (co), light hole (lh), heavy hole (hh) and split-off (so) bands.

Table 6.2: Energy gaps and effective masses in bulk for electrons (el), light holes (lh) and heavy holes (hh) at the Γ-point, unstressed and under biaxial tensile stress.

	$In_{0.53}Ga_{0.47}As$			$GaAs_{0.5}Sb_{0.5}$		
[eV]	0 MPa	500 MPa		0 MPa	500 MPa	
E_G	0.74	0.68		0.72	0.68	
Δ_{lh-hh}	0	0.04		0	0.04	
Δ_{so}	0.33	0.35		0.40	0.42	
$[m_e]$	x, y, z	x	y, z	x, y, z	x	y, z
m^*_{el}	0.041	0.037	0.040	0.046	0.041	0.040
m^*_{lh}	0.055	0.046	0.174	0.060	0.051	0.181
m^*_{hh}	0.450	0.452	0.068	0.455	0.458	0.074
m^*_{so}	0.135	0.154	0.121	0.155	0.172	0.142

The origin of the larger degeneracy is a strain-induced decrease in the DOS of the top of the valence band. The DOS is reduced both by a direction-dependent change in the band curvature and a lifting of the degeneracy of the heavy and light hole bands, as shown in Fig. 6.9. This is further illustrated in Table 6.2, which compares the effective masses and energy gaps at the Γ-point between the unstrained and the biaxial tensile case and shows a sharp decrease especially for the heavy hole mass in the y-and z-directions.

The higher currents at high V_{GS} are a consequence of a decrease in the tunneling bandgap compared to the unstrained case. Notice that in Fig. 6.9 this tunneling bandgap is formed between the first conduction band and the light hole band, which are coupled by the imaginary branches of the band structure. Because the light hole band shifts up under biaxial tensile strain, the tunneling bandgap decreases. Although the resulting current increase could be beneficial for the TFET I_{ON}, the SS degradation cancels out this advantage and severely limits the utility of stress in improving TFET performance.

6.2.3 Improved source design

The increase in source doping degeneracy due to strain can be counteracted with the improved source design introduced in Section 5.6.3, where it was applied to a pTFET. In the nTFETs studied here, the improved source with the lowly doped region results in a p^--p^+-n-i-n structure (see Fig. 6.6(b), from now on referred to as p-p-n-i-n TFET). The reduction in source doping degeneracy

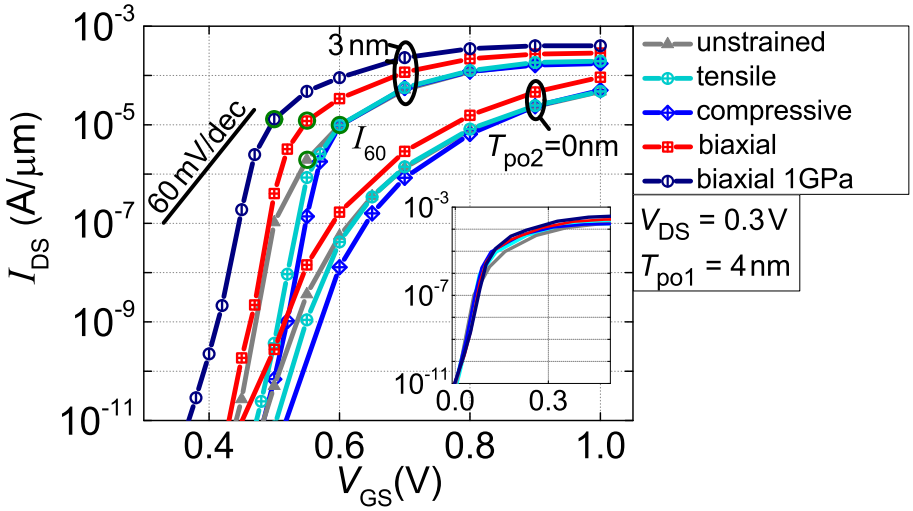

Figure 6.10: QM transfer characteristics of an unstressed $GaAs_{0.5}Sb_{0.5}$/ $In_{0.53}Ga_{0.47}As$ p-n-i-n TFET and uniformly stressed p-p-n-i-n hetero-TFETs for a T_{po2} of 3 nm and 0 nm. T_{po1} is 4 nm. Inset: 3 nm curves shifted to the same I_{OFF}. More configuration details in Fig. 6.6 and Table 6.1.

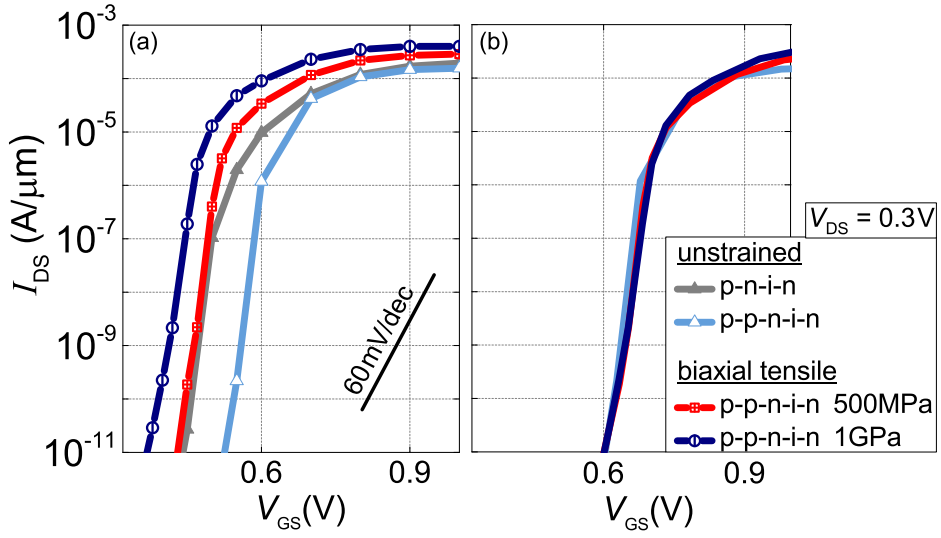

Figure 6.11: QM transfer characteristics of the 3 nm T_{po} p-n-i-n and associated p-p-n-i-n TFET of Fig. 6.10, the latter under increasing biaxial tensile stress. More configuration details in Fig. 6.6 and Table 6.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1x10^{-11}$A/μm coincides.

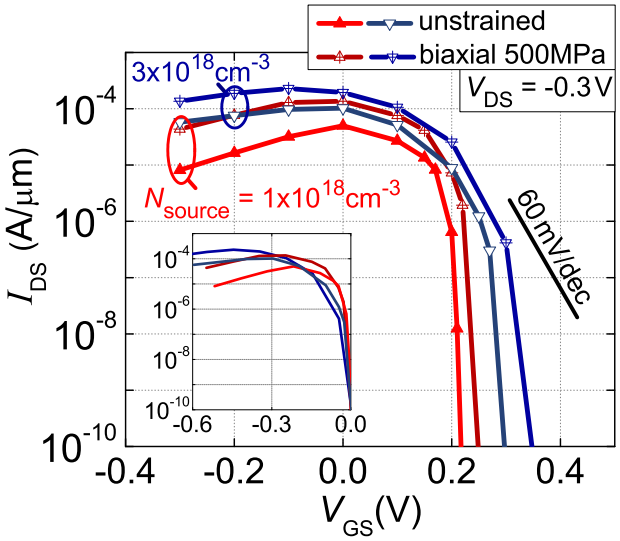

Figure 6.12: QM transfer characteristics of an unstressed and uniformly stressed n-n-p-i-p heterostructure $In_{0.53}Ga_{0.47}As/GaAs_{0.5}Sb_{0.5}$ TFETs for a T_{po1} of 3 nm and T_{po2} of 3 nm and two values for the source doping. Inset: The curves shifted to the same I_{OFF}. The configurations are the same as in Fig. 6.6(b), but with opposite doping types. More details in Table 6.1.

avoids the SS degradation, while maintaining improvements in I_{ON} thanks to the sharp band bending induced by the p^+-pocket. Fig. 6.10 shows that with the improved source design, an I_{60} of 1x10^1 μA/μm with an I_{ON} of 1.4x10^2 μA/μm is obtained for 500 MPa of biaxial tensile stress. With I_{OFF} fixed at 10^{-11} A/μm, I_{ON} is defined at V_{DD} = 0.3 V above the onset voltage. These performance metrics exceed the targets set in Section 2.6 to compete with MOSFET (I_{60} > 10 μA/μm and I_{ON} > 1x10^2 μA/μm for a V_{DD} < 0.5 V) and are similar to the results for pocketed heterostructure pointTFETs in Section 5.7.1, even though the source doping is lower. Increasing the stress to 1 GPa improves I_{ON} further to 1.7x10^2 μA/μm, with a similar I_{60} as for the 500 MPa case.

The improvement in I_{60} is mainly caused by the improved source design, while the improvement in I_{ON} is the result of the strain. We show this by simulating an identical unstrained p-p-n-i-n structure. Fig. 6.11 shows that this configuration has an I_{ON} of 1.1x10^2 μA/μm, while the shifted transfer characteristics coincide up to around 30 μA/μm with the strained case, which is above I_{60}. This is because the SS is determined mainly by the source degeneracy as the tunnel paths for these configurations are typically only a couple of nm (see e.g. Fig. 6.8).

With the improved source design, the degeneracy in both the unstrained and strained case is close to zero, resulting in a similar SS. The I_{ON} improvement in the strained case is limited, but in such a nearly optimal design this is to be expected. Nevertheless, these results show that we can use the improved source concept to benefit fully from the strain-induced I_{ON} improvement, while avoiding a degradation in SS. It also confirms the origin of the SS degradation to be the reduction of the DOS in the source region, rather than some other change in band structure.

For a pTFET, uniform strain does not degrade the SS as for an nTFET. The DOS in the pTFET source is determined by the conduction band, which is much less affected by the strain than the valence band DOS: the band shifts, but there is no degeneracy splitting as for the heavy hole band. Fig. 6.12 shows that applying a biaxial tensile stress of 500 MPa to an n-n-p-i-p TFET therefore directly improves I_{ON}, without an SS degradation. This is true for both a $1\text{x}10^{18}$ cm^{-3} and $3\text{x}10^{18}$ cm^{-3} source doping, although the former enjoys a larger relative benefit from the strain thanks to the lower source degeneracy. Both strained configurations obtain an I_{ON} of around $1.3\text{x}10^{2}$ μA/μm in a 0.3 V V_{DD} window. These simulations show that although the source DOS in a III-V pTFET is intrinsically low, requiring the application of the improved source concept, no additional measures are needed to benefit from the strain-induced I_{ON} improvement.

6.3 Non-uniform strain in lattice-mismatched heterostructure TFETs

In the previous section, strain was the result of a uniformly applied external stress. Alternatively, strain can also be inherently present at the heterojunction of two materials with a lattice mismatch. This mismatch (τ_0) is defined as the relative difference between the lattice constants of the two materials, and has a value typically limited to a few percent before relaxation through defect formation sets in. The strain profile generated by a mismatch at the tunnel junction is highly non-uniform. The first goal of this section is therefore to predict the complicated effect of such a non-uniform strain on the TFET performance. But next to simply predicting the TFET performance in the presence of an intrinsic non-uniform strain profile, we also want to assess whether it is possible to intentionally tune the lattice mismatch τ_0 to boost TFET performance over a lattice-matched combination. In particular, the localized nature of the strain profile provides the prospect to improve I_{ON} while preventing the SS degradation seen for uniformly strained configurations, since

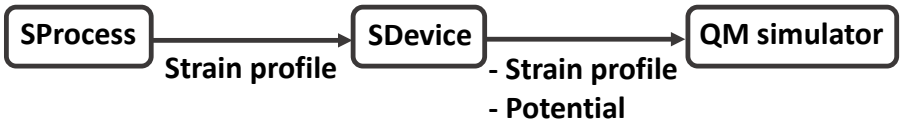

Figure 6.13: Simulation procedure for lattice-mismatched heterostructure TFETs with non-uniform strain profiles.

the source region remains unstrained. The strain could then be an alternative to solutions with dopant pockets, or could allow for a decrease in source doping.

6.3.1 Simulation details

We simulate heterostructure p-i-n TFETs in the $GaAs_xSb_{(1-x)}/In_yGa_{(1-y)}As$ system, which we have investigated in the previous chapters in the lattice-matched combination $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$. We will refer to this as the lattice-matched reference. By changing the alloy composition with x and y, a wide range of lattice-mismatched combinations are available. The TFET configurations we study are the same as in Fig. 6.6(a), but without a pocket.

Our simulation procedure of the lattice-mismatched configurations proceeds in several steps in different simulators (see Fig. 6.13). First, the strain profile is calculated in Synopsys Sentaurus Process, which solves the quasistatic equations of force equilibrium, taking into account the lattice mismatch [175]. In SProcess, the simulated structure is a slab of two different materials. The bottom material is unstrained and contains the source region. The top material is pseudomorphically grown on the first and contains the channel and drain regions. After deposition, the material stack is etched to the desired T_{body} of 10 nm and the top material is allowed to relax in the x and z-directions, regaining its bulk lattice constant away from the interface. No relaxation occurs in the y-direction, which is considered translationally invariant as before. After the relaxation, the source, drain and gate contacts are added. The non-uniform strain profile obtained after relaxation is then imported into SDevice, which calculates the electrostatic profile [51]. We verified that the strained bandgap profile as calculated by the deformation potential model in SDevice corresponds to our bulk thirty-band calculations. SDevice also adjusts the effective masses based on a $\mathbf{k}\cdot\mathbf{p}$-model. Finally, both the strain profile and the electrostatic potential are imported into our QM simulator, which constructs $\mathbf{H}_{\mathrm{S}}$ from the strain profile and calculates transport based on the electrostatic potential. $\mathbf{H}_{\mathrm{S}}$ for the non-uniform strain is constructed based on the position-dependent non-

diagonal expression in Eq. (6.8) and added to the spectral EF system as in Eq. (6.12).

To gain insight into the complicated effect of non-uniform strain on the TFET performance, we now proceed in two steps. In a first step, we assess the impact of additional non-uniform strain on the TFET performance. We start from a lattice-mismatched configuration with the same effective bandgap at the tunnel junction ($E_{\mathrm{G,eff}}$) before strain as the lattice-matched reference. We assess the impact of the additional strain and search for a lattice-mismatched material combination in the $\mathrm{GaAs_xSb_{(1-x)}/In_yGa_{(1-y)}As}$ system which outperforms the lattice-matched reference. In a second step, we investigate the intrinsic impact of the non-uniformity of the strain profile. In this case, we start from a lattice-mismatched configuration that has the same effective bandgap after the application of strain as the lattice-matched reference. This allows us to evaluate whether changes in TFET performance are due to a modification in $E_{\mathrm{G,eff}}$, or due to the non-uniform nature of the strain.

6.3.2 Additional impact of non-uniform strain

We first determine which sign of the mismatch τ_0 has a beneficial impact on the TFET performance by superimposing an artificial mismatch of five percent on the lattice-matched reference. All other parameters are kept the same. The transfer characteristics in Fig. 6.14 show that a negative mismatch, creating a tensile strain in the $\mathrm{In_{0.53}Ga_{0.47}As}$ regions, improves I_{ON}.

Next, we look for an existing material system with the same $E_{\mathrm{G,eff}}$ as the reference system before strain-induced band structure modifications are applied, such that we can study the "additional" impact of the non-uniform strain. To find this system, we need an expression as a function of composition for $E_{\mathrm{G,eff}}$ as well as τ_0. $E_{\mathrm{G,eff}}$ depends to first order on the electron affinities (E_{EA}) and E_{G} of the constituent materials (see Fig. 6.15), while τ_0, as discussed before, is given by the relative difference in lattice constants (L). Based on Fig. 6.15, we derive the following relation between $E_{\mathrm{G,eff}}$ and x and y for a general heterostucture system $\mathrm{A_xB_{(1-x)}/\ C_yD_{(1-y)}}$:

$$\begin{aligned} E_{\mathrm{G,eff}} &= E_{\mathrm{EA,1}} + E_{\mathrm{G,1}} - E_{\mathrm{EA,2}} \\ &= E_{\mathrm{EA,A}}\mathrm{x} + E_{\mathrm{EA,B}}(1\text{-}\mathrm{x}) + E_{\mathrm{G,A}}\mathrm{x} + E_{\mathrm{G,B}}(1\text{-}\mathrm{x}) \\ &\quad - W\mathrm{x}(1\text{-}\mathrm{x}) - E_{\mathrm{EA,C}}\mathrm{y} - E_{\mathrm{EA,D}}(1\text{-}\mathrm{y}) \end{aligned} \tag{6.18}$$

where we have used a linear interpolation for $E_{\mathrm{EA,1}}$ and $E_{\mathrm{EA,2}}$ and a non-linear interpolation with a bowing parameter W for $E_{\mathrm{G,1}}$ [139]. The lattice mismatch

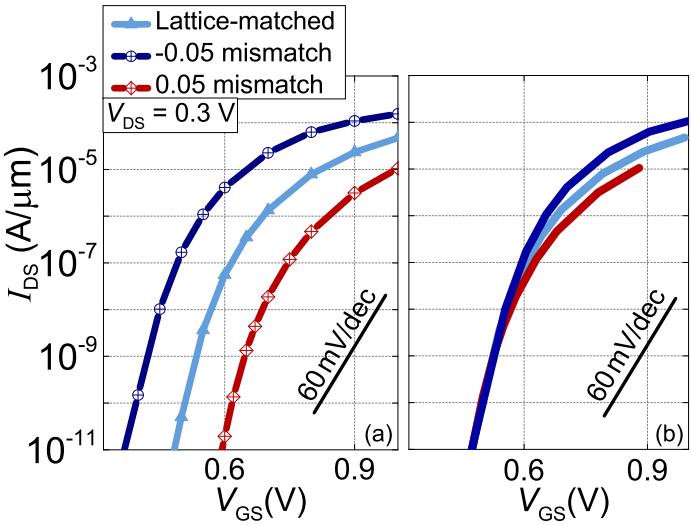

Figure 6.14: QM simulated transfer characteristics of a lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ TFET, compared with the same configuration superimposed with a positive and negative mismatch of five percent. T_{po} is 0 nm. More configuration details in Fig. 6.6(a) and Table 6.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1x10^{-11}$A/μm coincides.

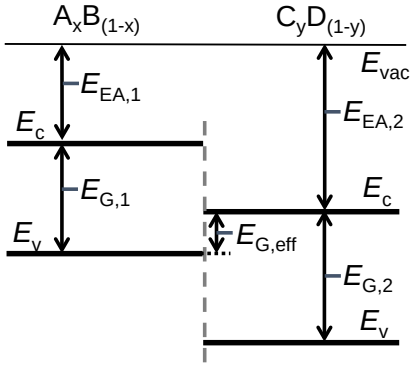

Figure 6.15: Band alignment at the heterojunction of a general $A_xB_{(1-x)}/C_yD_{(1-y)}$ system. The energy reference is the vacuum level E_{vac}.

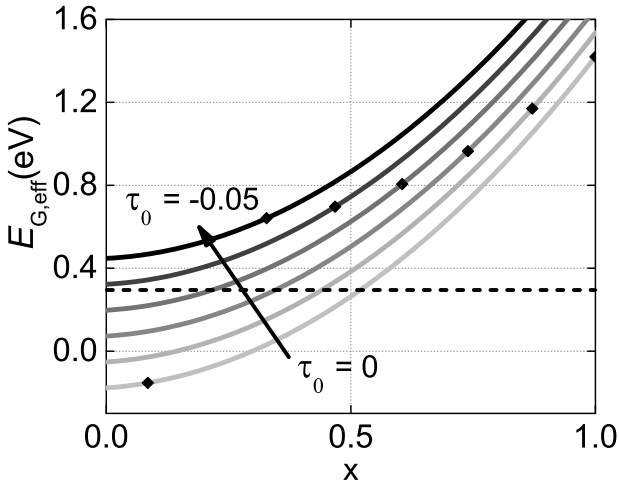

Figure 6.16: Effective bandgap based on unstrained bulk band alignment versus composition for the $GaAs_xSb_{(1-x)}/In_yGa_{(1-y)}As$ system according to Eq. (6.22) for different values of mismatch. The curves show the trade-off between $E_{G,eff}$ and τ_0 for a given x. The diamonds indicate the boundaries as in Eq. (6.21). The dashed line indicates $E_{G,eff}$ of the lattice-matched reference.

τ_0 can be written as a function of x and y as well:

$$\begin{aligned}\tau_0 &= \frac{L_2 - L_1}{L_1} \\ &= \frac{(L_C y + L_D(1\text{-}y)) - (L_A x + L_B(1\text{-}x))}{L_A x + L_B(1\text{-}x)}\end{aligned} \tag{6.19}$$

where we interpolate linearly for L_1 and L_2. From Eq. (6.19), we extract an expression for y as a function of x and τ_0:

$$y = \frac{(\tau_0 + 1)(L_A - L_B)x + (\tau_0 + 1)L_B - L_D}{L_C - L_D} \tag{6.20}$$

Inserting Eq. (6.20) into Eq. (6.18) results in a relation between $E_{G,eff}$, τ_0 and x. Note that the relationship between x and y in Eq. (6.20) puts extra boundaries

on x, since both x and y are limited to values between 0 and 1:

$$\mathrm{x} = \frac{L_\mathrm{D} - (\tau_0 + 1)L_\mathrm{B}}{(\tau_0 + 1)(L_\mathrm{A} - L_\mathrm{B})} \quad \text{for} \quad \mathrm{y} = 0$$

$$\mathrm{x} = \frac{L_\mathrm{C} - (\tau_0 + 1)L_\mathrm{B}}{(\tau_0 + 1)(L_\mathrm{A} - L_\mathrm{B})} \quad \text{for} \quad \mathrm{y} = 1 \tag{6.21}$$

Inserting the parameter values for the $\mathrm{GaAs_xSb_{(1-x)}/In_yGa_{(1-y)}As}$ system by taking A, B, C and D equal to GaAs, GaSb, InAs and GaAs respectively [176, 177, 178], we obtain for the combination of Eq. (6.18) and Eq. (6.20):

$$E_{\mathrm{G,eff}} = 1.43\mathrm{x}^2 + (0.91\tau_0 + 0.18)\mathrm{x} - 0.19 - 12.5\tau_0 \tag{6.22}$$

Fig. 6.16 shows the corresponding parabolae for different values of the mismatch τ_0. It is now clear that a trade-off in TFET performance exists between $E_{\mathrm{G,eff}}$ and τ_0 for a given value of x: a more beneficial (negative) τ_0 (see Fig. 6.14) results in a larger value for $E_{\mathrm{G,eff}}$, which degrades the TFET performance.

Next, we maximize the beneficial τ_0 with x constrained within the boundaries posed by Eq. (6.21) and with the added constraint that $E_{\mathrm{G,eff}}$ has the same value as the lattice-matched reference $\mathrm{GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As}$ (0.29 eV) such that we can study the added impact of strain. The resulting maximal mismatch is 3.7 percent, achieved for x = 0 and y = 0.53, which corresponds to a $\mathrm{GaSb/In_{0.53}Ga_{0.47}As}$ system.

We now compare a $\mathrm{GaSb/In_{0.53}Ga_{0.47}As}$ heterostructure TFET to the lattice-matched reference, which has the same $E_{\mathrm{G,eff}}$ before strain, to highlight the impact of the addition of non-uniform strain. We thereby assume that the $\mathrm{In_{0.53}Ga_{0.47}As}$ channel and drain are grown pseudomorphically on a relaxed GaSb source, as discussed in Section 6.3.1. The non-uniform profiles of the strain components around the tunnel junction as calculated by SProcess are shown in Fig. 6.17, with the resulting bandgap profile in Fig. 6.18(a). Both the body center and edge are heavily strained in the free x and z-directions, with an opposite sign in source and channel. The largest strain values are limited to a region close to the tunnel junction. Away from the junction, the structure has relaxed the stress in the x and z-directions to zero. A residual strain is still present in those two directions, however, as no relaxation occurs in the translationally invariant y-direction. The non-relaxed uniform stress in the y-direction of the channel and drain regions results in a strain in the x and z-direction in those regions. This residual strain is expected to disappear in a nanowire configuration, in which stress can relax in all three directions. These findings are in qualitative agreement with experimental studies on InAs/InP nanowires [179].

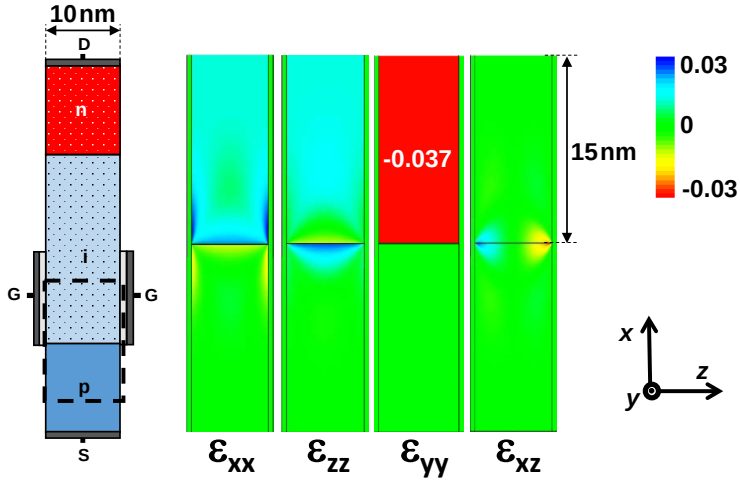

Figure 6.17: Profiles of the strain components in the boxed region around the tunnel junction as calculated by SProcess for a GaSb/$In_{0.53}Ga_{0.47}As$ configuration. The shear strain components that are not shown are negligible. The structure is considered translationally invariant in the y-direction.

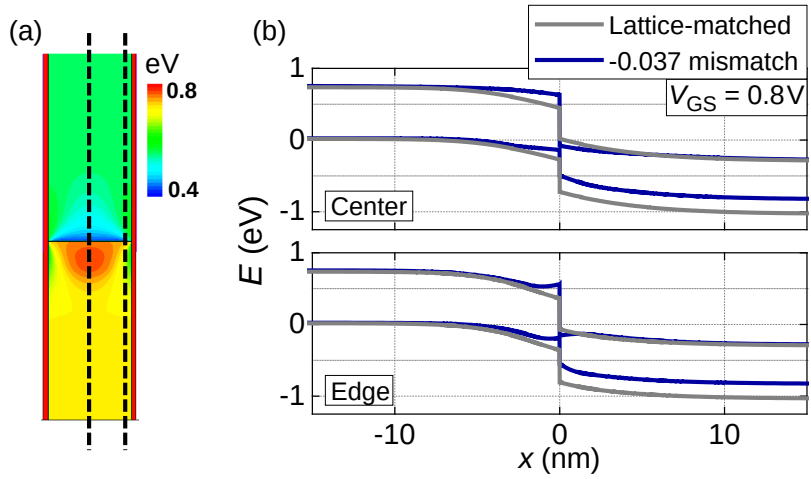

Figure 6.18: (a) Bandgap profile at the tunnel junction of a GaSb/$In_{0.53}Ga_{0.47}As$ TFET with the same dimensions as the profiles in Fig. 6.17. (b) Band diagrams along the cutlines indicated in (a) compared with those of an otherwise identical lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ configuration.

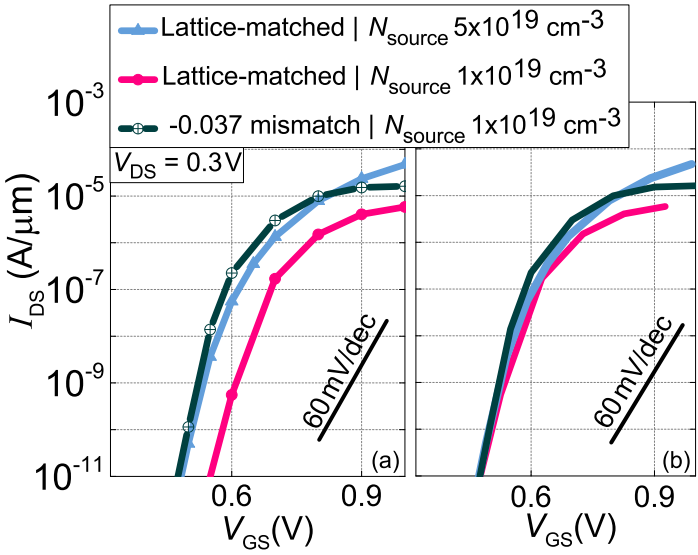

Figure 6.19: QM simulated transfer characteristics, comparing the lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ configuration with varying N_{source} to the lattice-mismatched $GaSb/In_{0.53}Ga_{0.47}As$ combination. T_{po} is 0 nm. More configuration details in Fig. 6.6(a) and Table 6.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is 1x10^{-11}A/μm coincides.

The transfer characteristics in Fig. 6.19 show an improvement of I_{ON} with a factor three in a V_{DD} window of 0.3 V for the lattice-mismatched configuration with a 1x10^{19} cm^{-3} source doping over a lattice-matched reference. The mismatched configuration with a 1x10^{19} cm^{-3} source doping attains even slightly better performance than the lattice-matched reference with a 5x10^{19} cm^{-3} source. Fig. 6.18(b) suggests that the improvement is a result of the reduced $E_{G,eff}$ at the tunnel junction both at the edge and in the center of the device, even though the bandgap locally increases at the source side in the center.

In contrast to a uniformly strained structure, the non-uniform strain in the source is limited to a small region around the junction, such that the DOS in the remainder of the source is unaffected. The heavy and light hole valence bands therefore do not split up away from the interface, retaining the same unstrained source doping degeneracy throughout most of the source region. This avoids the SS degradation seen in uniformly strained configurations and therefore obviates the need for an improved source design as in Section 6.2.3. Compared to the uniform case, the lattice mismatch does introduce the challenge of obtaining a defect-free strained interface.

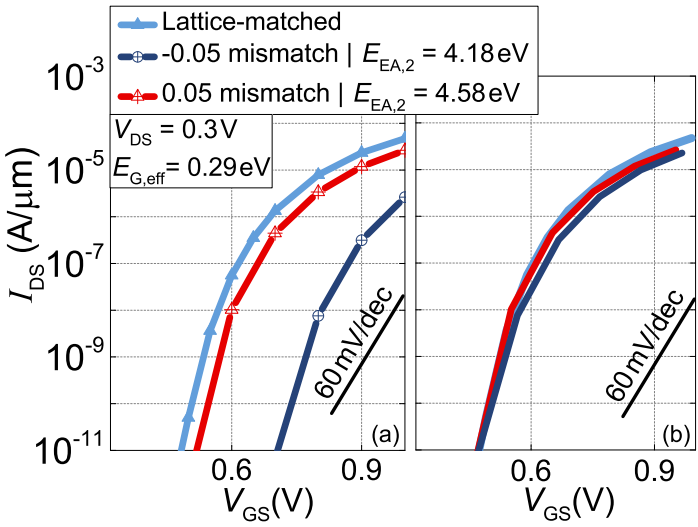

Figure 6.20: QM simulated transfer characteristics of a lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ configuration, compared with the same configuration superimposed with a positive and negative mismatch of five percent. The electron affinity of $In_{0.53}Ga_{0.47}As$ has been adjusted such that the average $E_{G,eff}$ after strain is 0.29 eV for all configurations. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1x10^{-11}A/\mu m$ coincides.

6.3.3 Intrinsic impact of non-uniform strain

We now take the opposite approach to the previous section: instead of choosing the material system such that $E_{G,eff}$ before strain-induced band modifications is the same as the lattice-matched reference, we look for a material system for which $E_{G,eff}$ *after* strain is the same as for the reference. We can then assess the specific effect of the non-uniformity of the strain profile on the TFET performance.

Unfortunately, no analytical expression is available for the variation of $E_{G,eff}$ with strain. We therefore take the lattice-matched reference $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ with a superimposed mismatch of five percent, as we did in Section 6.3.2, and artificially set $E_{EA,2}$ such that the average $E_{G,eff}$ after strain is equal to 0.29 eV, which is the unstrained value. Since $E_{G,eff}$ varies only weakly over the tunnel junction, taking the average has a small effect. The average $E_{G,eff}$ being equal in both TFETs, the only difference between the unstrained and the strained case is then caused by the non-uniformity of the strain profile

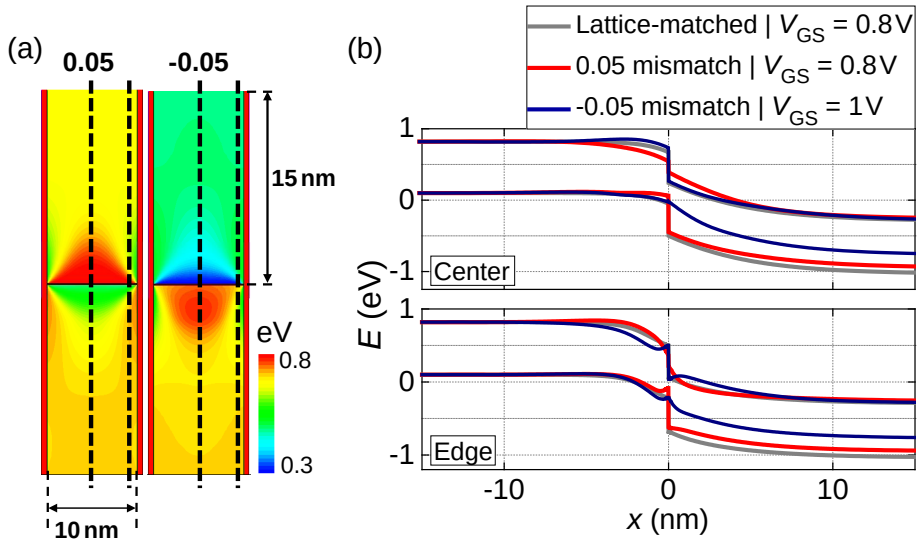

Figure 6.21: (a) Bandgap profile at the tunnel junction of the $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ TFETs in Fig. 6.20 with a superimposed positive and negative mismatch of five percent. (b) Band diagrams in the ON-state along the cutlines indicated in (a) compared with the lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ reference.

away from the junction.

Fig. 6.20 shows that the non-uniformity of the strain profile does not provide an additional advantage in SS or I_{ON}. The performance of the strained configurations with the same $E_{G,eff}$ after strain is similar to the unstrained case for the positive (+0.05) mismatch and slightly degraded for the negative (-0.05) mismatch. In the band diagrams of Fig. 6.21 and the bandgap and electron affinity cutlines of Fig. 6.22, we see that both in the center and at the edge of the device, a decrease in bandgap at one side of the tunnel junction is compensated by a bandgap increase at the other side. Although the shifts in the band edges of the individual materials shift the onset voltage of the TFET by 0.22 V (see Fig. 6.20), their net impact on the tunnel path length is limited at a similar voltage beyond BTBT onset (see Fig. 6.21(b)), which is reflected in the similar transfer characteristics of Fig. 6.20.

Note that it is possible that for a pTFET, a non-uniform strain profile exists that is beneficial for the output characteristics, because the local splitting of the heavy and light hole valence band lowers the DOS in the channel region [180]. The exploration of this large design space falls outside of the scope of

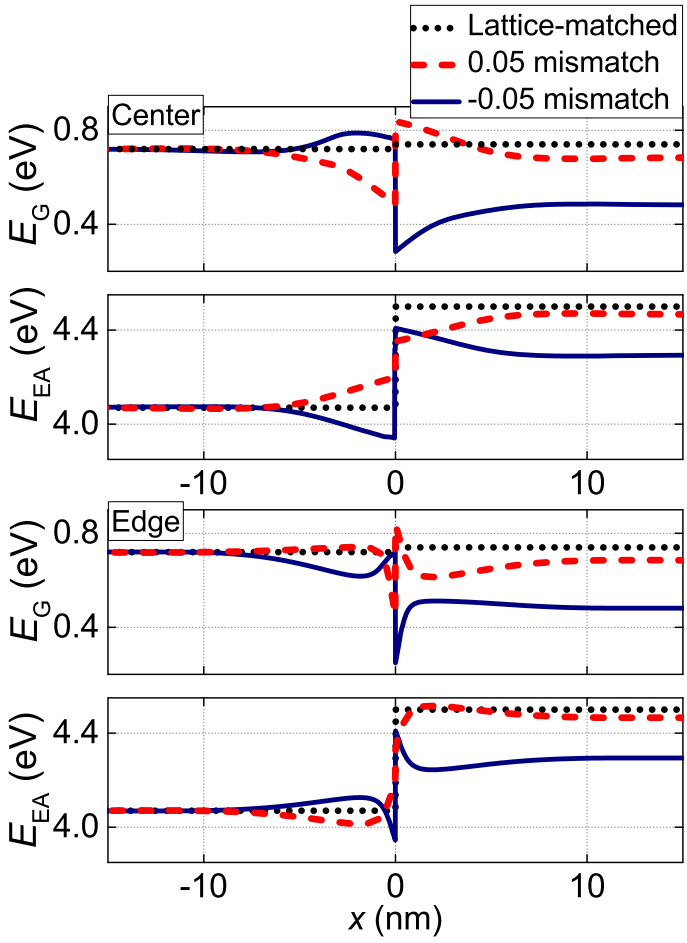

Figure 6.22: Bandgap and electron affinity along the cutlines indicated in Fig. 6.21(a) compared with the relaxed $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ reference.

this work.

6.4 Conclusion

In this chapter, we included the effects of spin and mechanical strain in our EF formalism. The inclusion of spin doubles the fifteen-band basis set to a thirty-band model. The main effect on the band structure is a splitting of the degenerate valence bands due to spin-orbit coupling. We showed that the resulting change in DOS has a significant effect on the tunneling current for configurations with a T_{body} below 10 nm. Strain was included with a position-dependent eight-band strain Hamiltonian, valid for low k-values.

With the thirty-band implementation, we first investigated uniform stress applied to lattice-matched $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ heterostructure TFETs. We verified the degradation of SS for nTFETs under uniaxial or biaxial stress reported in literature, and traced its origin to a strain-induced reduction in valence band DOS. We were able to avoid the SS degradation by applying the improved source design with a lowly doped source region introduced in Chapter 5. This allowed us to take full benefit of the I_{ON} improvement under biaxial tensile stress.

In a next step, we studied non-uniform strain profiles, caused by a lattice mismatch at the heterojunction. We found that for a properly chosen material system, the strain at the junction reduces $E_{\mathrm{G,eff}}$, thereby increasing I_{ON}. The localization of the strain profile around the junction also prevents the SS degradation found in the case of uniform stress. We checked with an artificial lattice-mismatched configuration with the same $E_{\mathrm{G,eff}}$ as the lattice-matched reference that it is indeed the $E_{\mathrm{G,eff}}$ reduction which is responsible for the performance improvement, not the shape of the non-uniform strain profile. In conclusion, the lattice mismatch is an extra parameter which extends the design space of heterostructure TFETs. It should be kept in mind, however, that experimentally, the lattice mismatch is limited by the capability of obtaining defect-free strained interfaces.

The results on the uniform strain in this chapter have been published in *Electron Device Letters* [109]. Part of the results on the non-uniform strain have been reported at the *European Solid-State Device Research Conference (ESSDERC)* [181].

Chapter 7

Self-consistency

In the previous chapters, the simulations to determine the diode and TFET performance were not self-consistent. In particular, the electrostatic potential for the QM simulations was extracted from a SC calculation in SDevice. The effect of quantum phenomena, like SIQC, on the electrostatic potential is thereby not taken into account. SIQC, for instance, results in a redistribution of the charge density and a modification of the band structure, both of which influence the electrostatic potential. For TFET structures with confined dimensions (below 10 nm), we hence expect a significant change in band bending and doping degeneracy compared to wider architectures, which will impact the TFET performance. In this chapter, we therefore develop a self-consistent procedure which connects the charge density, calculated from the solutions of the EF system, with the electrostatic potential, calculated with Poisson's equation. The aim is to open up the application of our formalism to TFET configurations with a small body thickness. This will allow us to determine below which body thickness self-consistent simulations become indispensable and to do a first assessment of the impact of confinement on the results of the previous chapters.

The chapter is structured as follows. We first show how to calculate the carrier density from properly normalized EFs (Section 7.1). Next, we use the carrier density to update the electrostatic potential with Poisson's equation and Neumann boundary conditions at the open contacts (Section 7.2). We find that a straightforward loop of Poisson's equation and the carrier density calculation is not stable. We therefore employ a Gummel scheme, combined with successive underrelaxation to obtain self-consistency of the electrostatic potential and the carrier density (Section 7.3). Finally, we illustrate the effect of

the self-consistency on both weakly and strongly confined TFET configurations by comparing with non-self-consistent simulations (Section 7.4). The results of this chapter are summarized in the conclusion (Section 7.5).

7.1 Carrier density

To start off the self-consistent procedure, we first calculate the carrier density from a set of EFs, obtained from the solution of an EF system with an arbitrary number of bands. In this section, we first make sure that the corresponding state wave function is properly normalized (Section 7.1.1). Normalization was not required in Chapter 3, where we calculated transmission probabilities, because any normalization constant would cancel out. This is not the case here, where we want to calculate a carrier density. After determining the normalization constant, we derive an expression for the carrier density (Section 7.1.2).

7.1.1 Normalization

In the EF formalism, the state wave functions are decomposed into a set of EFs and basis functions (see Eq. (3.4)). In the contacts, these EFs consist of modes that are injected, reflected and transmitted in the device (see Eqs. (3.17) and (3.18)). One state then corresponds to a single injected mode, combined with all of its associated reflected modes in the same contact and transmitted modes in the other contact. We identify a state with three quantum numbers: the values of k_x, k_y and the index γ of the injected subband mode. This yields the following form for the wave function of a single state:

$$\psi_{k_x,k_y,\gamma}(\mathbf{r}) = \begin{cases} Ae^{ik_y y}\sum_n^N \left[I_\gamma e^{ik_x x}\chi_{n\gamma}(z) + \sum_{\gamma'} r_{\gamma'} e^{ik_{x,\gamma'}x}\chi_{n\gamma'}(z)\right] U_n(\mathbf{r}) & x \in \text{inj} \\ Ae^{ik_y y}\sum_n^N \left[\sum_\delta t_\delta e^{ik_{x,\delta}x}\chi_{n\delta}(z)\right] U_n(\mathbf{r}) & x \in \text{tr} \end{cases}$$

$$= Ae^{ik_y y}\sum_n^N \left[\sum_\zeta \sum_l \left[c_{n\zeta l}(x)e^{ik_{x,\zeta,l}x}\right]\chi_{n\zeta}(z)\right] U_n(\mathbf{r}) \tag{7.1}$$

where A is the normalization constant of the full state and "inj" and "tr" denote the contact of injection and transmission respectively. The index n runs over all considered bands N as before. The index ζ runs over all distinct subband numbers. Some modes have the same subband number and therefore share the same confined envelope $\chi_{n\zeta}(z)$, e.g. an injected mode at k_x and its

corresponding reflected mode on the same subband at $-k_x$. We count such modes, which correspond to the same ζ but have a different value of k_x, with the index l. The definition of ζ ensures the orthogonality of $\chi_{n\zeta}$ and $\chi_{n\zeta'}$ if $\zeta \neq \zeta'$, a property which we will need later on. $c_{n\zeta l}(x)$ denotes the coefficients of the different modes, in which any mode and band dependent scaling factors have been absorbed. These coefficients have a position dependence like a step function: they are only non-zero in the contact associated to the mode. For simplicity, we will drop the x-dependence of the $c_{n\zeta l}$ in the notation.

To normalize the states in Eq. (7.1), we combine a box normalization to a Kronecker delta in the confined z-direction with Dirac delta normalization in the infinite x- and y-directions. The latter is required because the plane waves in x and y would otherwise render the normalization integral infinite [182, 183]. The normalization condition for the state wave functions of Eq. (7.1) is then:

$$\int_{\mathbb{R}^3} d\mathbf{r}\, \psi^*_{k'_x,k'_y,\gamma'}(\mathbf{r})\psi_{k_x,k_y,\gamma}(\mathbf{r}) = \delta(k_x - k'_x)\delta(k_y - k'_y)\delta_{\gamma\gamma'} \tag{7.2}$$

with $\delta_{\gamma\gamma'}$ a Kronecker delta. Since the normalization integral runs over all space, from $-\infty$ to $+\infty$, we can neglect the contribution of the active region and normalize the states in the half-infinite contacts.

We now derive an expression for A, such that the normalization condition Eq. (7.2) is satisfied. We start by inserting the expression for ψ from Eq. (7.1) into Eq. (7.2). This yields:

$$\begin{aligned}
&\int_{\mathbb{R}^3} d\mathbf{r}\, \psi^*_{k'_x,k'_y,\gamma'}(\mathbf{r})\psi_{k_x,k_y,\gamma}(\mathbf{r}) \\
&= \int_{\mathbb{R}^3} d\mathbf{r} \left[A^* e^{-ik'_y y} \sum_{m,\zeta',l'} c^*_{m\zeta'l'} e^{-ik'_{x,\zeta',l'}x} \chi^*_{m\zeta'}(z) U^*_m(\mathbf{r}) \right] \\
&\qquad \times \left[A e^{ik_y y} \sum_{n,\zeta,l} c_{n\zeta l} e^{ik_{x,\zeta,l}x} \chi_{n\zeta}(z) U_n(\mathbf{r}) \right] \\
&= |A|^2 \int_{\mathbb{R}^3} d\mathbf{r}\, e^{i(k_y - k'_y)y} \sum_{\substack{\zeta',\zeta \\ l',l}} e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} \sum_{n,m} c^*_{m\zeta'l'} c_{n\zeta l} \chi^*_{m\zeta'}(z)\chi_{n\zeta}(z) U^*_m(\mathbf{r}) U_n(\mathbf{r})
\end{aligned} \tag{7.3}$$

To get rid of the basis functions, we convert the integral over the entire position space to a sum of integrals over the unit cell:

$$\int_{\mathbb{R}^3} d\mathbf{r} = \sum_{N_{\mathrm{uc}}} \int_{V_{\mathrm{uc}}} d\mathbf{r} \tag{7.4}$$

where V_{uc} is the volume of a unit cell and N_{uc} is the number of unit cells in the crystal under consideration. Over the volume of a unit cell, the basis functions by definition oscillate much faster than the EFs (see Section 3.2.1). We therefore consider the latter to be constant over V_{uc}, allowing us to take all associated factors out of the integral. Eq. (7.3) can then be rewritten as:

$$|A|^2 \sum_{N_{\mathrm{uc}}} e^{i(k_y - k'_y)y} \sum_{\substack{\zeta',\zeta \\ l',l}} e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} \sum_{n,m} c^*_{m\zeta'l'} c_{n\zeta l} \chi^*_{m\zeta'}(z) \chi_{n\zeta}(z) \int_{V_{\mathrm{uc}}} d\mathbf{r}\, U^*_m(\mathbf{r}) U_n(\mathbf{r}) \tag{7.5}$$

The basis functions are defined to be normalized to the unit cell volume, such that:

$$\int_{V_{\mathrm{uc}}} d\mathbf{r}\, U^*_m(\mathbf{r}) U_n(\mathbf{r}) = V_{\mathrm{uc}}\, \delta_{nm} \tag{7.6}$$

Inserting this relation into Eq. (7.5), we obtain:

$$|A|^2 \sum_{N_{\mathrm{uc}}} e^{i(k_y - k'_y)y} \sum_{\substack{\zeta',\zeta \\ l',l}} e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} \sum_{n} c^*_{n\zeta'l'} c_{n\zeta l} \chi^*_{n\zeta'}(z) \chi_{n\zeta}(z)\, V_{\mathrm{uc}} \tag{7.7}$$

We can then consider the sum over N_{uc} with the factor V_{uc} as a Riemann sum, and convert it back to an integral over the entire crystal volume:

$$|A|^2 \int_{\mathbb{R}^3} d\mathbf{r}\, e^{i(k_y - k'_y)y} \sum_{\substack{\zeta',\zeta \\ l',l}} e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} \sum_{n} c^*_{n\zeta'l'} c_{n\zeta l} \chi^*_{n\zeta'}(z) \chi_{n\zeta}(z) \tag{7.8}$$

We convert the plane wave factors in x and y into Dirac delta functions, applying the following properties:

$$\int_{-\infty}^{\infty} dy\, e^{i(k_y - k'_y)y} = 2\pi\delta(k_y - k'_y)$$

$$\int_{-\infty}^{0} dx\, e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} = \int_{L_D}^{\infty} dx\, e^{i(k_{x,\zeta,l} - k'_{x,\zeta',l'})x} = \pi\delta(k_{x,\zeta,l} - k'_{x,\zeta',l'}) \tag{7.9}$$

where 0 and L_D are the left and right boundaries respectively of the device active area and the integral in x runs from $-\infty$ to 0 for modes in the left contact, and from L_D to $+\infty$ for modes in the right contact. We then obtain for Eq. (7.8):

$$|A|^2 2\pi\,\delta(k_y - k'_y) \sum_{\substack{\zeta',\zeta \\ l',l}} \pi\delta(k_{x,\zeta,l} - k'_{x,\zeta',l'}) \int dz \sum_n c^*_{n\zeta' l'} c_{n\zeta l} \chi^*_{n\zeta'}(z) \chi_{n\zeta}(z) \quad (7.10)$$

We know that the $\chi_{n\zeta}$ are orthogonal, such that for $\zeta \neq \zeta'$ Eq. (7.10) is zero. For $l \neq l'$, $k_{x,\zeta,l} - k'_{x,\zeta',l'}$ cannot be zero by definition of l: l counts modes with the same subband number ζ, but a distinct k_x-value. The corresponding delta function $\delta(k_{x,\zeta,l} - k'_{x,\zeta',l'})$ therefore must be zero. Consequently, the only non-zero case is when $\zeta = \zeta'$ and $l = l'$:

$$|A|^2 2\pi\,\delta(k_y - k'_y) \sum_{\zeta,l} \pi\,\delta(k_{x,\zeta,l} - k'_{x,\zeta,l}) \int dz \sum_n |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2 \quad (7.11)$$

To obtain a form similar to the right-hand side of Eq. (7.2), we now want to take the delta function out of the summation over ζ and l. We therefore refer $\delta(k_{x,\zeta,l} - k'_{x,\zeta,l})$ to the wave number k_x of the injected mode. Since each $k_{x,\zeta,l}$ is linked to a specific k_x (see Eq. (7.1)), we can define a function $g(k_x) = k_{x,\zeta,l}(k_x) - k_{x,\zeta,l}(k'_x)$ and make use of this property of the delta function:

$$\delta(g(x)) = \sum_i \frac{\delta(x - x_i)}{|g'(x_i)|} \quad (7.12)$$

where x_i are the zeroes of $g(x)$. Since in our case $g(k_x)$ has only one zero at $k_x = k'_x$, we obtain:

$$\delta(k_{x,\zeta,l}(k_x) - k_{x,\zeta,l}(k'_x)) = \frac{\delta(k_x - k'_x)}{\left|\frac{d\,k_{x,\zeta,l}}{dk_x}\right|_{k'_x}} \quad (7.13)$$

Inserting Eq. (7.13) into Eq. (7.11) yields:

$$|A|^2 2\pi^2 \delta(k_y - k'_y) \sum_{\zeta,l} \frac{\delta(k_x - k'_x)}{\left|\frac{d\,k_{x,\zeta,l}}{dk_x}\right|_{k'_x}} \int dz \sum_n |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2$$

$$= |A|^2 2\pi^2 \delta(k_y - k'_y)\,\delta(k_x - k'_x) \sum_{\zeta,l} \left|\frac{d\,k_{x,\zeta,l}}{dk_x}\right|^{-1}_{k'_x} \int dz \sum_n |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2 \quad (7.14)$$

Comparing Eq. (7.14) to the normalization condition in Eq. (7.2), we obtain the following condition for the state wave functions to be normalized:

$$|A|^2 2\pi^2 \sum_{\zeta,l} \left| \frac{d\,k_{x,\zeta,l}}{dk_x} \right|_{k'_x}^{-1} \int dz \sum_n |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2 = 1 \tag{7.15}$$

which is satisfied if:

$$|A| = \frac{1}{\pi \sqrt{2 \sum_{\zeta,l} \int dz \sum_n \left| \frac{d\,k_{x,\zeta,l}}{dk_x} \right|_{k'_x}^{-1} |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2}} \tag{7.16}$$

The factor $\left| \frac{d\,k_{x,\zeta,l}}{dk_x} \right|_{k'_x}$ incorporates the differences in local slope, and hence DOS, between the reference injected subband and the other subbands. This is visualized in Fig. 7.1: for a finite shift Δk_x of the injected mode, the reflected modes shift with $\Delta k_{x,1}$ and $\Delta k_{x,2}$ and the transmitted mode with $\Delta k_{x,3}$. The size of these shifts is determined by the local curvature of the subbands. The factor $\left| \frac{d\,k_{x,\zeta,l}}{dk_x} \right|_{k'_x}$ is then the relative size of the shifts in the limit of Δk_x going to zero. Note that once the EFs have been normalized in the contacts, they are normalized in the entire device region, as no scattering is present.

Instead of k_x, E can also be chosen as a quantum number. The normalization condition of Eq. (7.2) is then modified to:

$$\int_{\mathbb{R}^3} d\mathbf{r}\, \psi^*_{E',k'_y,\gamma'}(\mathbf{r}) \psi_{E,k_y,\gamma}(\mathbf{r}) = \delta(E - E')\delta(k_y - k'_y)\delta_{\gamma\gamma'} \tag{7.17}$$

Following a derivation analogous to the one that led to Eq. (7.16), we obtain for the normalization constant:

$$|A| = \frac{1}{\pi \sqrt{2 \sum_{\zeta,l} \int dz \sum_n \left| \frac{d\,k_{x,\zeta,l}}{dE} \right|_{E'}^{-1} |c_{n\zeta l}|^2 |\chi_{n\zeta}(z)|^2}} \tag{7.18}$$

where now the factor $\left| \frac{d\,k_{x,\zeta,l}}{dE} \right|_{E'}$ connects the normalization to the DOS.

In the remainder of this chapter, we assume the normalization is carried out with k_x as the quantum number, because we will see in the next section that it results in a current density formula that is easier to implement.

7.1.2 Calculating carrier density

With the state wave functions normalized, we now derive a formula to calculate the total carrier density in the device based on the EFs. We start from the

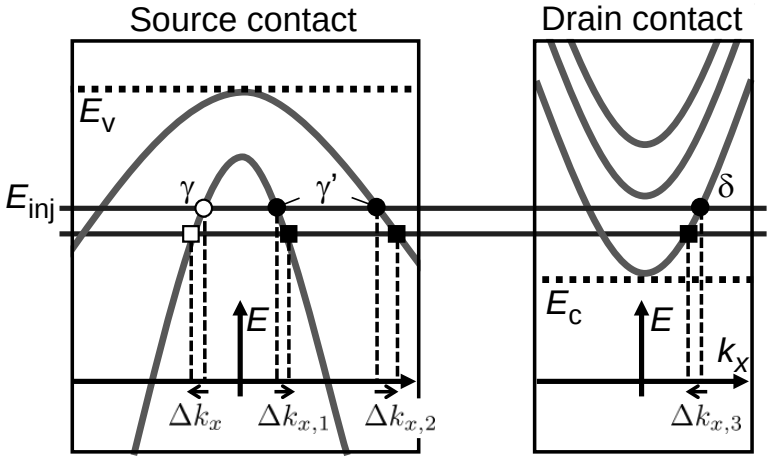

Figure 7.1: Schematic subband structure in the source and drain contact. A single mode is injected (open circle) into the device at an injection energy E_{inj}. The probability current is transmitted to one mode at the other contact and reflected to two modes at the injection contact (closed circles). We then shift the injected mode with Δk_x. The Greek symbols correspond to the respective mode indices.

general expression for total free carrier density: a sum for the discrete quantum number γ and an integration for the continuous quantum numbers k_x and k_y of the normalized probability densities of all injected states (in both contacts), weighted with the appropriate Fermi-Dirac distribution:

$$p(\mathbf{r}) + n(\mathbf{r}) = \int_{k_x} dk_x \int_{k_y} dk_y \sum_{\gamma} \psi^*_{k_x,k_y,\gamma}(\mathbf{r}) \psi_{k_x,k_y,\gamma}(\mathbf{r}) f_c(E(k_x,k_y,\gamma)) \quad (7.19)$$

where n is the electron density and p is the hole density. f_c contains the carrier Fermi-Dirac distribution of electrons or holes, depending on the energy E:

$$f_c(E(k_x,k_y,\gamma)) = \begin{cases} f(E(k_x,k_y,\gamma) - E_{\mathrm{F}}) & E > E_{\mathrm{c}} \quad \text{(electrons)} \\ f(E_{\mathrm{F}} - E(k_x,k_y,\gamma)) & E < E_{\mathrm{v}} \quad \text{(holes)} \end{cases} \quad (7.20)$$

with $f(E_\Delta)$ defined as $1/(e^{E_\Delta/kT}+1)$ and E_{c}, E_{v} and E_{F} the band edges and quasi-Fermi level at the contact at which the state $\psi_{k_x,k_y,\gamma}$ is injected. This is consistent with the Landauer formalism we assumed in Section 3.5, in which the occupation of each state is determined throughout the device by its contact of injection [184]. In the case of BTBT, this means an electron state is weighted as an electron state also after the tunneling event. The same is true for a hole

state. In the remainder of the derivation, we will suppress the dependency of E on k_x, k_y, γ in the notation.

Inserting the EF expansion of the state wave functions into Eq. (7.19) yields:

$$\int_{k_x} dk_x \int_{k_y} dk_y \sum_{\gamma} \left[\sum_{m} F^*_{m,k_x,k_y,\gamma}(\mathbf{r}) U^*_m(\mathbf{r}) \right] \left[\sum_{n} F_{n,k_x,k_y,\gamma}(\mathbf{r}) U_n(\mathbf{r}) \right] f_c(E)$$

$$= \int_{k_x} dk_x \int_{k_y} dk_y \sum_{n,m,\gamma} F^*_{m,k_x,k_y,\gamma}(\mathbf{r}) F_{n,k_x,k_y,\gamma}(\mathbf{r}) U^*_m(\mathbf{r}) U_n(\mathbf{r}) f_c(E) \tag{7.21}$$

where m and n are understood to run over all considered bands N as before. Eq. (7.21) still contains the basis functions, which are not calculated in the solution of the EF system. To obtain an expression based solely on the EFs, we average the density in Eq. (7.21) over a unit cell and use the same assumption as in Eq. (7.5), namely that the EFs can be considered constant over a unit cell:

$$\frac{1}{V_{\mathrm{uc}}} \int_{V_{\mathrm{uc}}} d\mathbf{r} \int_{k_x} dk_x \int_{k_y} dk_y \sum_{n,m,\gamma} F^*_{m,k_x,k_y,\gamma}(\mathbf{r}) F_{n,k_x,k_y,\gamma}(\mathbf{r}) U^*_m(\mathbf{r}) U_n(\mathbf{r}) f_c(E)$$

$$= \int_{k_x} dk_x \int_{k_y} dk_y \sum_{n,m,\gamma} F^*_{m,k_x,k_y,\gamma}(\mathbf{r}) F_{n,k_x,k_y,\gamma}(\mathbf{r}) \frac{1}{V_{\mathrm{uc}}} \int_{V_{\mathrm{uc}}} d\mathbf{r}\, U^*_m(\mathbf{r}) U_n(\mathbf{r}) f_c(E)$$

$$= \int_{k_x} dk_x \int_{k_y} dk_y \sum_{\gamma,n} F^*_{n,k_x,k_y,\gamma}(\mathbf{r}) F_{n,k_x,k_y,\gamma}(\mathbf{r}) f_c(E)$$

$$= \int_{k_x} dk_x \int_{k_y} dk_y \sum_{\gamma,n} |F_{n,k_x,k_y,\gamma}(\mathbf{r})|^2 f_c(E) \tag{7.22}$$

where we have used the normalization of the basis functions to the unit cell volume Eq. (7.6). For our 2D case, Eq. (7.22) reduces to:

$$p(\mathbf{r}) + n(\mathbf{r}) = \int_{k_x} dk_x \int_{k_y} dk_y \sum_{\gamma,n} |F_{n,k_x,k_y,\gamma}(x,z)|^2 f_c(E) \tag{7.23}$$

where we assign the carrier density to $p(\mathbf{r})$ for $E < E_{\mathrm{v}}$ and to $n(\mathbf{r})$ if $E > E_{\mathrm{c}}$ at the contact of injection.

Note that the carrier density in Eq. (7.23) is the result of a ballistic calculation. This means the carriers do not rethermalize anywhere in the device (an issue we also encountered in Section 5.6.3). Quantum wells that have a potential below the injection window therefore remain unfilled. Such a situation can occur e.g. in an nTFET if the gate voltage is so high to push the energy bands in the

channel below those in the drain. Care has been taken in the remainder of the chapter to avoid such a quantum well.

Eq. (7.23) now provides a way to calculate the free carrier density based on the calculated EFs. The next step is then to update the electrostatic potential accordingly.

7.2 Potential

Once we have found the free carrier density profile in the device, the electrostatic potential can be updated with Poisson's equation. This constitutes a single iteration of the self-consistent loop. In this section, we first discuss the general form of Poisson's equation (Section 7.2.1), and then determine the appropriate boundary conditions for its solution (Section 7.2.2).

7.2.1 Poisson's equation

Poisson's equation connects the electrostatic potential in the device with the present charge density:

$$\nabla^2\phi(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon} = -\frac{q(p(\mathbf{r}) - n(\mathbf{r}) + N_\mathrm{D}(\mathbf{r}) - N_\mathrm{A}(\mathbf{r}))}{\epsilon} \tag{7.24}$$

with ϕ the electrostatic potential, ϵ the permittivity, also known as the dielectric constant, ρ the total charge density and N_A and N_D respectively the acceptor and donor dopant concentrations. The dopants are assumed to be fully ionized throughout the device. This approximation becomes questionable for strongly confined structures, if the bandgap widening results in a larger ionization energy. To solve Eq. (7.24), we discretize n, p, N_A, N_D and ϕ on the spatial mesh and use a FD scheme for the Laplacian as we did in Section 3.4 for the EF system. The resulting discretized system has dimensions $N_x \cdot N_z \times N_x \cdot N_z$. It is therefore considerably smaller than the EF system, which is a factor N larger in each dimension. To solve the Poisson system, we now have to specify the boundary conditions.

7.2.2 Boundary conditions

We start with the boundary conditions in the confined z-direction. In contrast to the hard wall boundary conditions applied to the EF system, we do include the gate dielectric in the solution of Poisson's equation. The gate voltage is

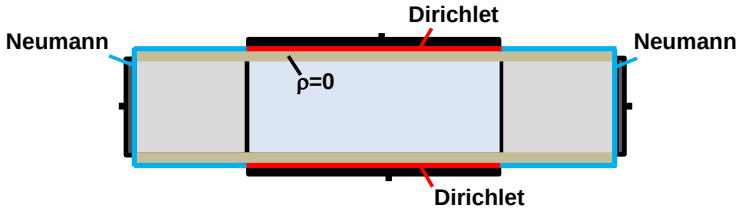

Figure 7.2: Boundary conditions for the solution of Poisson's equation for a generic TFET structure. The boundary conditions are applied at the outer edge of the gate dielectric.

then applied as a Dirichlet (fixed-value) boundary condition directly to that part of the outside boundary of the dielectric which is covered by the gate (see Fig. 7.2). The actual applied electrostatic potential boundary value is:

$$\phi_{\text{gate}} = V_{\text{GS}} - \phi_{\text{ms}} \tag{7.25}$$

where ϕ_{ms} is the work function difference between the gate metal and the semiconductor underneath the dielectric. The part of the boundary in the confined z-direction not covered by the gate is assumed to be electric field-free, which corresponds to Neumann boundary conditions. For the bulk of the dielectric, we impose the net charge to be zero. The inclusion of the gate dielectric in this way is advantageous for two reasons. First, the drop of the gate potential in the dielectric is automatically included in the solution of Poisson's equation. Second, because of the absence of any net charge, the potential can settle inside this region, avoiding abrupt changes and ensuring a smooth potential at the device boundaries. In practice, the dielectric is implemented as an extension of the device semiconductor with an equivalent thickness, calculated based on the dielectric constants.

For the open source and drain contacts in the x-direction, no single established set of boundary conditions exists in literature. Both Dirichlet and Neumann boundary conditions are applied in various schemes [185, 186, 187]. For Dirichlet boundary conditions, the applied voltage fixes the quasi-Fermi levels at source and drain, while the values for the electrostatic potential are obtained from a 1D charge neutrality calculation at each contact. When simulating transport, however, the fixed Dirichlet boundaries can lead to charge accumulation (when reflection is high) or depletion (when transmission is high) near the contacts. This complicates the convergence of the self-consistent loop. To remedy the charge imbalance, two strategies are generally employed. One is to adjust the quasi-Fermi levels at the contacts in each iteration until charge neutrality is reached either locally at each contact separately or globally for all contacts at the same time. A second strategy is to assign the injected states with a drifted

Fermi-level such that either charge neutrality or current continuity is obtained [185]. A comparison of these methods for a resonant tunneling diode was made by Potz [186]. The disadvantage is that both strategies require an additional loop to ensure either charge neutrality or current conservation.

We therefore choose to impose Neumann boundary conditions at the source and drain contacts, which do not give rise to charge accumulation/depletion issues [187]. This entails the assumption that the electric field goes to zero in the transport direction at the source and drain contact. In our application, this is not an additional assumption, as it is already inherent to QTBM, which we use for the boundary conditions of the EF system (see Section 3.3). The Neumann conditions ensure a flat potential at the contacts, while allowing the potential to "float" up or down in subsequent iterations of the self-consistent loop to accommodate for an accumulation or depletion of charge. In the final converged solution, charge neutrality is achieved globally at all contacts. A disadvantage is that the Neumann boundary conditions themselves only specify the derivative at the boundaries: the obtained potential is therefore only defined up to a constant. For TFETs, this constant is fixed by the Dirichlet conditions at the gate. For diodes, we either have to specify the value of the potential at a single point or impose an additional constraint, e.g. that the mean of the potential over the device is zero. A summary of the implemented boundary conditions for Poisson's equation is presented in Fig. 7.2.

We now have the necessary building blocks for a single iteration of the self-consistent loop: the calculation of the carrier density and the corresponding update of the electrostatic potential. The next step is then to repeat this process: the carrier density is recalculated based on the new potential, after which an updated potential is determined and so on. To ensure that this loop is stable, we have to develop a self-consistency strategy.

7.3 Self-consistency

We found that simply looping between the calculation of the charge density with Eq. (7.23) and the electrostatic potential with Eq. (7.24), leads to instabilities. The new potential at each iteration overcompensates for the lack of charge neutrality, leading to strong oscillations. We therefore need to include some kind of damping between the different iterations. This is typically done by modifying the electrostatic potential at the end of each iteration, before the calculation of a new charge density. In this section, we discuss several convergence schemes (Section 7.3.1) and select a strategy for our application (Section 7.3.2).

7.3.1 Convergence schemes

Three prominent schemes are typically used for Schrödinger-Poisson type loops: Newton(-Broyden) [185, 188], Gummel [189] and successive underrelaxation (SUR) [190]. We briefly list pros (+) and cons (−) of these methods found in literature:

- Newton(-Broyden)
  - \+ Fast convergence, even when equations are strongly coupled
  - − Requires good initial guess
  - − Requires calculation of Jacobian: computationally intensive (although approximative schemes exist)

- Gummel
  - \+ Low computational cost
  - \+ Fast initial error reduction
  - \+ Weighted, position-dependent mixing of old and new electrostatic potentials based on carrier densities
  - − Slow convergence if equations are strongly coupled (e.g. for large source-drain or gate-source bias)

- Successive underrelaxation
  - \+ Low computational cost
  - \+ Easy to understand and implement
  - \+ Direct control over damping
  - − Damping constant not based on carrier densities
  - − Crude uniform mixing

Because of the additional computational requirements to calculate a Jacobian in the Newton method, we opt for the Gummel scheme, supplemented with SUR if additional damping is required.

7.3.2 Gummel + SUR

In the Gummel iteration scheme, a term is added to both sides of Poisson's equation in Eq. (7.24) that mixes the new potential, ϕ_{new}, with the potential of the previous iteration, ϕ_{old}, to dampen abrupt changes between iterations [189]. Eq. (7.24) is then modified to:

$$\nabla^2\phi_{\text{new}}(\mathbf{r}) - \frac{q(p(\mathbf{r}) + n(\mathbf{r}))}{\epsilon V_{\text{ref}}}\phi_{\text{new}}(\mathbf{r}) = -\frac{\rho(\mathbf{r})}{\epsilon} - \frac{q(p(\mathbf{r}) + n(\mathbf{r}))}{\epsilon V_{\text{ref}}}\phi_{\text{old}}(\mathbf{r}) \quad (7.26)$$

with V_{ref} the reference voltage, which determines the damping and hence also the convergence speed of the self-consistent loop. A larger value of V_{ref} corresponds to a smaller damping and a faster convergence, but also a larger risk of divergence. A typical value is the thermal voltage $\frac{kT}{q}$, which we use as the default. If exact convergence would be achieved, $\phi_{\text{new}}(\mathbf{r}) = \phi_{\text{old}}(\mathbf{r})$, and Eq. (7.26) is reduced to Poisson's equation of Eq. (7.24). Convergence can be monitored by checking the norm of the relative error in the potential after each iteration:

$$\Delta\phi = \left\| \frac{\boldsymbol{\phi}_{\text{new}} - \boldsymbol{\phi}_{\text{old}}}{\boldsymbol{\phi}_{\text{old}}} \right\|_{\infty} \tag{7.27}$$

where the infinity norm is defined as the maximum of the absolute values of the elements, and where $\boldsymbol{\phi}_{\text{new}}$ and $\boldsymbol{\phi}_{\text{old}}$ are discretized vectors of ϕ_{new} and ϕ_{old} respectively over the full spatial mesh. In practice, convergence is not exact, but is determined to be achieved when the residual $\Delta\phi$ goes below a certain threshold (typically taken around $1\text{x}10^{-4}$).

The amount of iterations to achieve convergence depends strongly on the initial guess for ϕ_{old}. The closer it is to the final converged result, the fewer iterations are typically required. One option for the initial guess is a simple constant potential, or a potential that varies linearly from source to drain. These examples take almost no time to construct, but result in a long process of convergence. We therefore opt for a more accurate initial guess, namely the SC potential calculated by SDevice. This takes longer to construct, but the extra time is greatly outweighed by the reduction in the number of iterations in the self-consistent loop. When simulating multiple bias points, the converged potential of the previous bias point is taken as initial guess. This further decreases the amount of iterations (typically around 10).

In cases where quantum effects are important, the SC initial potential guess might still be quite different from the final converged result, resulting in a large overshoot in the first few iterations. This can be detected from a sudden increase in the residual $\Delta\phi$ (typically if $\Delta\phi > 1$). If overshoot occurs, usually in the first iteration, we damp more strongly by reducing V_{ref}. In addition, we supplement the Gummel scheme with SUR for additional damping. In SUR, the potential calculated from Eq. (7.26) is mixed with the potential of the previous iteration [190]:

$$\phi_{\text{new}}(\mathbf{r}) = \omega\phi_{\text{new}}(\mathbf{r}) + (1-\omega)\phi_{\text{old}}(\mathbf{r}) \tag{7.28}$$

with ω a damping parameter, for which we take a starting value of 1 (no damping). If overshoot is detected, ω is progressively reduced. The full procedure is summarized in the flowchart of Fig. 7.3.

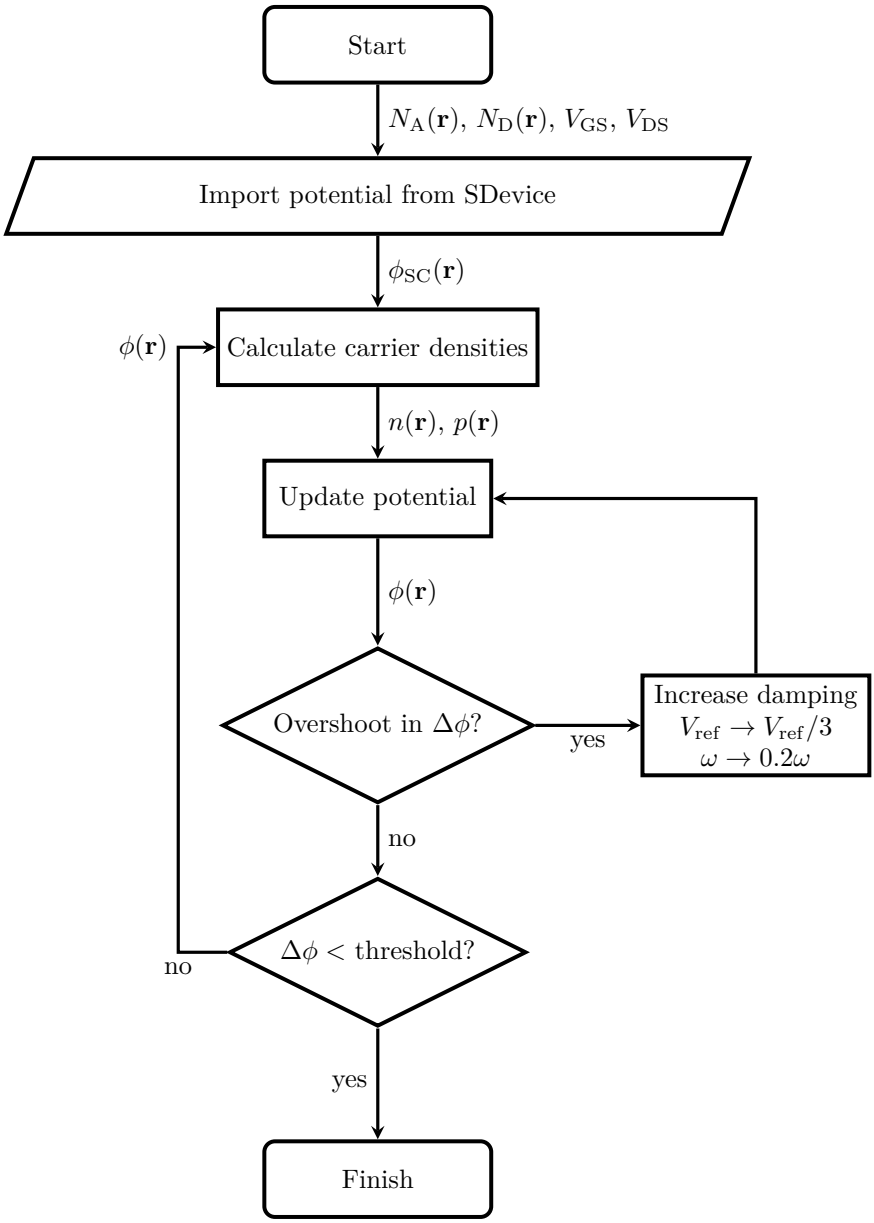

Figure 7.3: Full self-consistent simulation procedure. ϕ_{SC} is the initial semiclassical guess for the potential.

7.4 Self-consistent versus non-self-consistent simulations

With the self-consistent procedure implemented, we can now compare with non-self-consistent simulations and assess as an example the impact of the SIQC-induced redistribution of carriers on the electrostatic potential and transfer characteristics for different degrees of confinement.

7.4.1 Simulation details

The simulated configurations are homostructure $In_{0.53}Ga_{0.47}As$ pocketed pointTFETs, with a T_{po} of 3 nm and a T_{body} of 10 nm and 5 nm. For the latter, we expect a larger effect of SIQC on the electrostatic potential and hence the transfer characteristics. The TFET configurations correspond to the ones discussed in Section 5.5.4, and are shown in Fig. 5.8(a), with details in Table 5.1. The source and drain regions have been extended to 40 nm and 35 nm respectively to help with convergence. The drain doping is also chosen higher at $1x10^{19}$ cm^{-3}.

The QM simulations are carried out with the fifteen-band model of Chapter 5. The fifteen-band parameters can be found in Appendix C.2.1. The SC potential

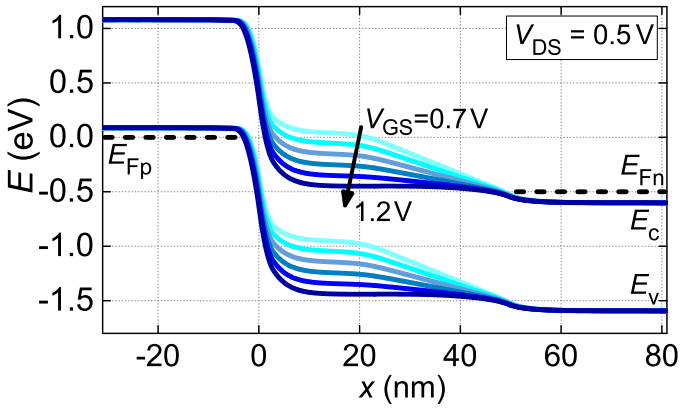

Figure 7.4: Self-consistent confined band diagrams through the center of the device of an $In_{0.53}Ga_{0.47}As$ pocketed pointTFET with a T_{body} of 5 nm and a T_{po} of 3 nm for V_{GS} increasing in steps of 0.1 V from 0.7 V to 1.2 V. More details on the configurations can be found in Fig. 5.8(a) and Table 5.1.

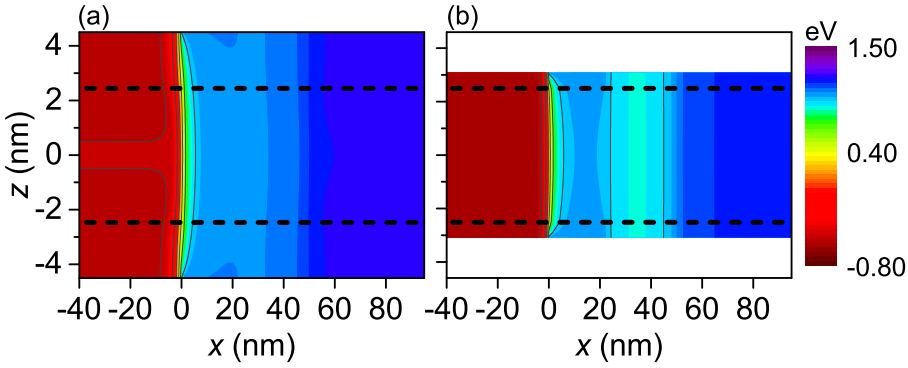

Figure 7.5: Comparison of (a) self-consistent and (b) non-self-consistent 2D electrostatic potential of an $In_{0.53}Ga_{0.47}As$ pocketed pointTFET with a T_{body} of 5 nm and a T_{po} of 3 nm for a V_{GS} of 1.2 V. The horizontal dashed lines indicate the boundaries between the gate dielectric and the semiconductor. In the self-consistent simulation, the dielectric is implemented as a layer of $In_{0.53}Ga_{0.47}As$ with an equivalent thickness. More details on the configurations can be found in Fig. 5.8(a) and Table 5.1.

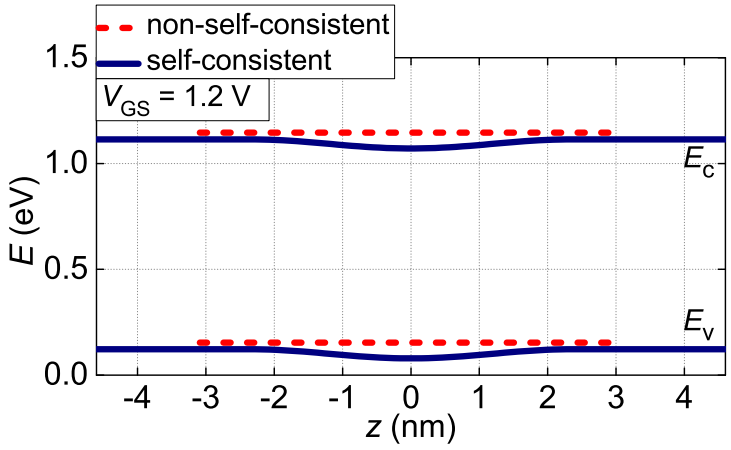

Figure 7.6: Comparison of self-consistent and non-self-consistent confined band diagrams along a vertical cutlines at $x = -40$ nm in Fig. 7.5.

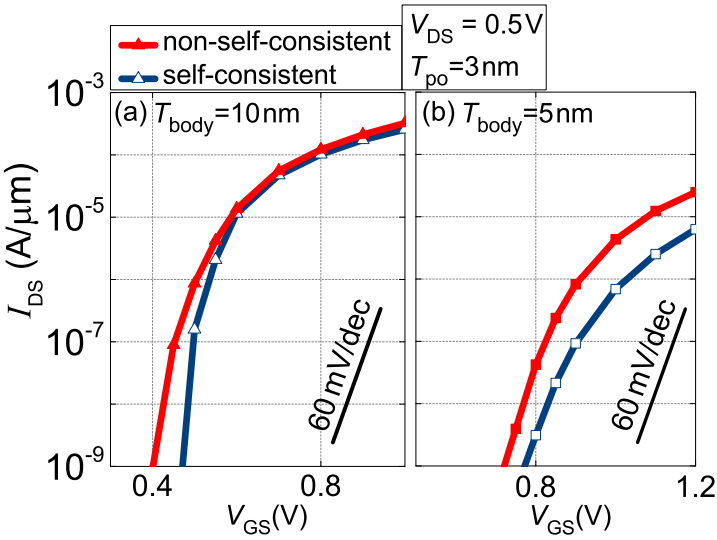

Figure 7.7: QM simulated transfer characteristics comparing self-consistent to non-self-consistent simulations of $In_{0.53}Ga_{0.47}As$ pocketed pointTFETs with two values for T_{body}. More details on the configurations can be found in Fig. 5.8(a) and Table 5.1.

which is used for the non-self-consistent QM simulations and which serves as the initial guess for the self-consistent loop is calculated with SDevice. The SC parameters are listed in Appendix D.3.2.

The calculation of the carrier density requires a finer energy mesh than the transport calculations. Compared to the transmission probabilities, the carrier density has a more erratic behavior, with peaks around band edges and at possible resonant energy levels. We therefore increase the number of energy points from 40 to 500. The points are chosen adaptively around the subband edges in the contacts, based on a decreasing derivative in the E-k relation and weighted for importance with the appropriate Fermi-Dirac distribution at each contact. The energy mesh is additionally refined around (resonance) peaks in the transmission spectrum.

7.4.2 Results

As a check of proper convergence, we first plot the band diagrams of the converged solutions along a cutline through the center of the device for different

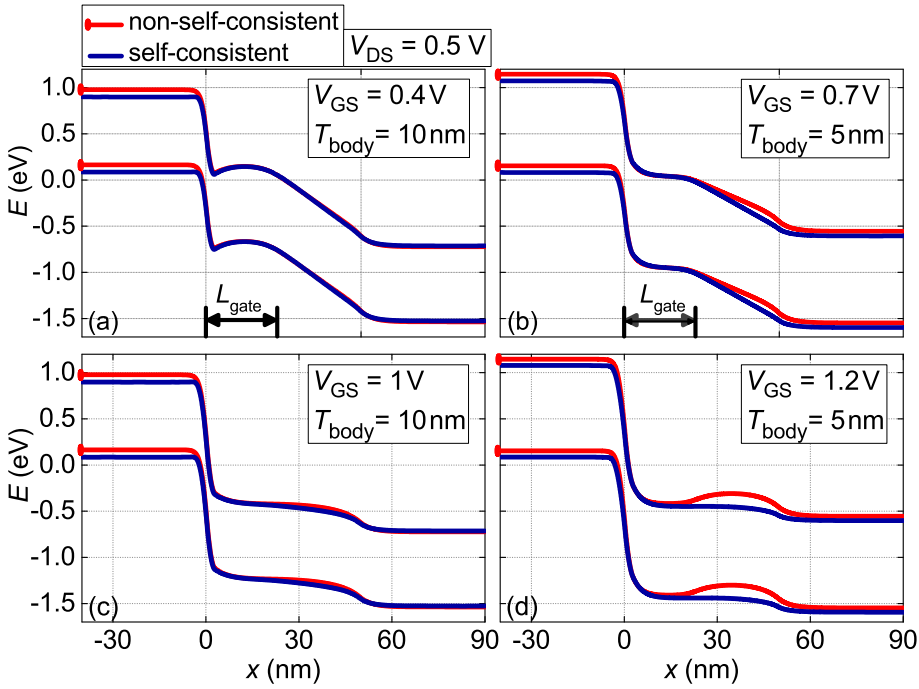

Figure 7.8: Self-consistent and non-self-consistent confined band diagrams through the center of the devices of Fig. 7.7 at onset and in the ON-state.

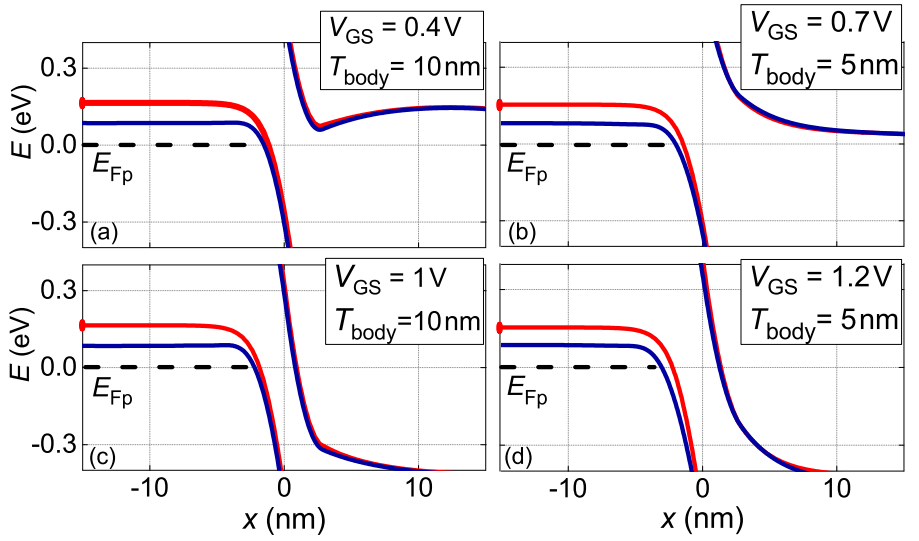

Figure 7.9: Zoom of the band diagrams in Fig. 7.8 around the tunnel junction.

values of V_{GS}. Fig. 7.4 shows for the 5 nm T_{body} configuration that the bands of the converged solutions are flat in the contact regions as imposed by the Neumann boundary conditions of Poisson's equation. The bands have also converged to the same value in the contact regions for all values of V_{GS}, indicating that the electrostatic potential is properly converged for each bias point.

For the highest V_{GS}, we compare self-consistent and non-self-consistent 2D potential profiles in Fig. 7.5 and band diagrams along a vertical cutline in the source in Fig. 7.6. Notice that in the source region of the self-consistent case, the bands are slightly curved because of the confinement of the carriers to the center of the device (Fig. 7.6).

Fig. 7.7 then shows the difference between self-consistent and non-self-consistent transfer characteristics of the pocketed pointTFETs for two values of T_{body}. As expected, the discrepancy is largest for the most confined configuration of 5 nm, while the 10 nm results only diverge for lower values of V_{GS}.

The band diagrams in Figs. 7.8(a) and (b) and 7.9(a) and (b) show that the origin of the discrepancy at low V_{GS} between self-consistent and non-self-consistent transfer characteristics lies mainly with a difference in source doping degeneracy. In the non-self-consistent simulations, the Fermi level is positioned such that the bulk free carrier density, calculated based on a set of bands with a simple non-parabolic correction, compensates for the dopant charge. In self-consistent simulations, on the other hand, the free carrier density is calculated based on the full confined fifteen-band band structure. This confined band structure has a higher valence band DOS than the effective mass based SC bulk DOS, hence the smaller degeneracy in the self-consistent case. Also at the drain side, the degeneracy is smaller for the self-consistent case, especially for the 10 nm T_{body}.

The difference in discrepancy at high V_{GS} is more difficult to explain from the 1D cutlines in this single example. Figs. 7.8(c) and (d) and 7.9(c) and (d) show that in the center of the device, there is a larger discrepancy in electric field at the tunnel junction for the 5 nm configuration than for the 10 nm configuration. In the 10 nm case, the tunnel path lengths are similar below E_{Fp}. In the ON-state, however, the tunnel paths at the edges of the device are most important, and they have a similar length in the 5 nm and 10 nm configurations. The full explanation can therefore not entirely be inferred from 1D cutlines, but is rather an effect of the full 2D potential.

7.5 Conclusion

In this chapter, we developed a self-consistent procedure for the calculation of the QM carrier density and electrostatic potential. We derived a normalization for the EFs, based on Dirac delta normalization in the transport direction and box normalization in the confined direction. We then derived a formula to calculate carrier density based on the EFs. With this density, we updated the electrostatic potential by solving the Gummel version of Poisson's equation with Dirichlet boundary conditions at the gate contact and Neumann boundary conditions elsewhere, including at the open source and drain contacts. We complemented the Gummel scheme with SUR to ensure sufficient damping and prevent overshoots in the self-consistent loop.

The self-consistent procedure opened up the application domain of our formalism to strongly confined devices. We showed stable convergence for fifteen-band simulations of $In_{0.53}Ga_{0.47}As$ homostructure pocketed n-channel pointTFET and found that the transfer characteristics of self-consistent and non-self-consistent simulations diverge significantly for a confined T_{body} of 5 nm, with the self-consistent simulations of this n-channel device predicting generally smaller tunneling currents. The band diagrams showed the origin to be a smaller source doping degeneracy for the self-consistent case, combined with a corresponding smaller electric field. For a T_{body} of 10 nm, the discrepancy was found to be smaller, and mainly limited to small values of V_{GS} due to a difference in source doping degeneracy.

These results show that the non-self-consistent simulations can be used for a qualitative study, as long as T_{body} is sufficiently large (not below 10 nm for $In_{0.53}Ga_{0.47}As$).

Chapter 8

Electron-phonon scattering in resonant TFETs

The TFET simulations in the previous chapters are all ballistic. In actual devices, however, carriers are subjected to a number of scattering mechanisms due to phonons, impurities, roughness and others, which have the potential to degrade device performance. In this chapter, we specifically focus on electron-phonon scattering and investigate the performance sensitivity of two TFET configurations: a conventional heterostructure TFET and a resonant TFET. The latter is a configuration which depends on resonant tunneling to improve transmission and which greatly outperforms conventional heterostructure TFETs in ballistic simulations. However, because of its dependence on coherent transport, the resonant TFET is expected to show a larger sensitivity to scattering. Our aim is therefore to assess whether the resonant TFET still outperforms the conventional TFET in the presence of electron-phonon scattering. This chapter differs from the other chapters in that it does not rely on the QM simulation approach developed in the previous chapters. The simulations were carried out with the NanoElectronics MOdeling 5 package (NEMO5) during a research stay at Purdue University, IN, USA.

The chapter is structured as follows. We first briefly introduce the general formalism used in our NEMO5 simulations, which is based on the tight-binding (TB) band structure model and the non-equilibrium Green's function (NEGF) method (Section 8.1). We then discuss the scattering self-energies for the different types of included phonons (Section 8.2). Next, we introduce the resonant TFET and compare the sensitivity of its performance to electron-phonon scattering to a conventional heterostructure TFET (Section 8.3).

Finally, we estimate to which bulk mobilities the simulated scattering strengths correspond (Section 8.4). The chapter ends with the conclusions (Section 8.5).

8.1 General formalism

We simulate the effect of electron-phonon scattering on TFET performance with NEMO5, which is developed and maintained at Purdue University [191]. NEMO5 is a fully QM simulator, based on the NEGF formalism. In this section, we discuss the chosen TB band structure model (Section 8.1.1) and the advantages of the NEGF approach for scattering simulations (Section 8.1.2). We then briefly outline the simulation procedure followed by NEMO5 (Section 8.1.3). More detailed information on NEMO5 can be found in the referenced publications.

8.1.1 Band structure model

For the band structure model, NEMO5 offers a choice between various TB band structure models and an eight-band $\mathbf{k}\cdot\mathbf{p}$-model. We choose the full-zone $sp^3d^5s^*$ TB basis set with spin-orbit coupling, because it is the most complete model available. In its description of the first Brillouin zone, it corresponds to our thirty-band EF model. The atomistic TB basis has the advantage of retaining its validity for very confined structures (even below 2 nm), for which the continuum approach of $\mathbf{k}\cdot\mathbf{p}$/EF-methods fails. On the other hand, as the location of the atoms determines the numerical solution grid, the computational demand quickly becomes unfeasible for bigger structures.

8.1.2 Solution method

NEMO5 is based on the NEGF approach (briefly discussed in Section 2.5.2), which is more suited to model scattering than our own WF-based EF formalism. Our WF approach is a single-particle method: the WF that is calculated describes a single particle moving through the device. To include scattering between (quasi-)particles in this WF approach, we would have to modify the EF system to a many-body equation. In NEGF, on the other hand, the solution functions, the Green's functions, are correlation functions, which means they do not describe a single particle, but rather its probability in moving from one point to the other. Scattering can therefore be incorporated as a modification of this probability. Each scattering mechanism corresponds to a self-energy as explained in Section 2.5.2, which is derived from many-body theory.

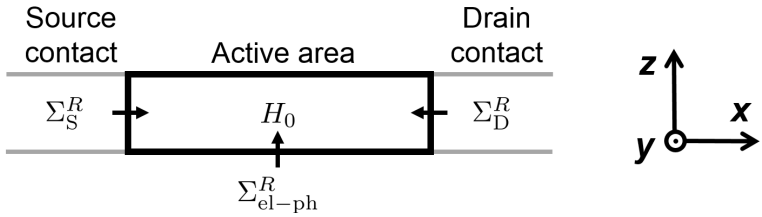

Figure 8.1: Schematic device configuration, illustrating the different terms in the NEGF system.

8.1.3 Simulation procedure

We now summarize the simulation procedure followed by NEMO5. First, NEMO5 constructs the NEGF system, which consists of the Hamiltonian of the closed system H_0 and the self-energies in the chosen basis (see also Fig. 8.1). The system is constructed as in Eq. (2.22), reformulated in a TB basis as:

$$(E - H_0(\alpha,\beta) - \Sigma^R(\alpha,\beta,E))G^R(\alpha,\beta,E) = 1 \tag{8.1}$$

or equivalently:

$$(E - H_0(\alpha,\beta) - \Sigma^R(\alpha,\beta,E))^{-1} = G^R(\alpha,\beta,E) \tag{8.2}$$

with the index α denoting the TB orbitals at the atom β and Σ^R the retarded self-energy:

$$\Sigma^R(\alpha,\beta,E) = \Sigma^R_{\mathrm{S/D}}(\alpha,\beta,E) + \Sigma^R_{\mathrm{el-ph}}(\alpha,\beta,E) \tag{8.3}$$

which has a contribution from the open source and drain contacts $\Sigma^R_{\mathrm{S/D}}$ and the electron-phonon interaction $\Sigma^R_{\mathrm{el-ph}}$. The contact self-energy is determined with the transfer matrix method [19, 192]. The electron-phonon self-energy is the sum of a self-energy for the acoustic phonons, Σ^R_{ac}, and one for the optical phonons, Σ^R_{opt}. The expressions of these self-energies are discussed in Section 8.2. The next step is the inversion of the matrix on the left-hand side of Eq. (8.2). To reduce the computational burden, the inversion is carried out numerically with the RGF method [56]. The result is the retarded Green's function, from which NEMO5 then calculates the lesser Green's function $G^<$ and greater Green's function $G^>$:

$$G^{<,>}(\alpha,\beta,E) = \int d\alpha' d\beta' G^R(\alpha,\alpha',E)\Sigma^{<,>}(\alpha',\beta',E)G^{R\dagger}(\beta',\beta,E) \tag{8.4}$$

with the Hermitian conjugate of G^R defined as:

$$G^{R\dagger}(\beta',\beta,E) = \overline{G}^R(\beta,\beta',E) \tag{8.5}$$

$\Sigma^{<,>}$ is the lesser or greater self-energy, which has a contact contribution and a electron-phonon contribution analogous to Σ^R:

$$\Sigma^{<,>}(\alpha,\beta,E) = \Sigma^{<,>}_{\mathrm{S/D}}(\alpha,\beta,E) + \Sigma^{<,>}_{\mathrm{el-ph}}(\alpha,\beta,E) \tag{8.6}$$

in which $\Sigma^{<,>}_{\mathrm{el-ph}}$ is again the sum of a lesser or greater self-energy for the acoustic phonons, $\Sigma^{<,>}_{\mathrm{ac}}$, and for the optical phonons, $\Sigma^{<,>}_{\mathrm{opt}}$. The expressions for these self-energies are discussed in Section 8.2 as well. From $G^<$, the electron density at each atom position is calculated as:

$$n(\mathbf{r}) = \int \frac{dE}{2\pi}\,\mathrm{Im}\left\{\sum_{\alpha_{\mathrm{at}}} G^<(\alpha_{\mathrm{at}},\alpha_{\mathrm{at}},E)\right\} \tag{8.7}$$

where the index α_{at} runs over the orbitals associated to the atom at position $\mathbf{r}$. The hole density is analogously calculated from $G^>$:

$$p(\mathbf{r}) = \int \frac{dE}{2\pi}\,\mathrm{Im}\left\{\sum_{\alpha_{\mathrm{at}}} G^>(\alpha_{\mathrm{at}},\alpha_{\mathrm{at}},E)\right\} \tag{8.8}$$

The electron and hole carrier densities are then inserted into the Poisson equation as in Eq. (7.24) to update the electrostatic potential. This forms a self-consistent loop. Once the carrier densities and electrostatic potential have converged, current between two atom layers i and i+1 in the transport direction is given by [193]:

$$J_{\mathrm{i}\to\mathrm{i+1}} = \frac{q}{\hbar}\int \frac{dE}{2\pi} 2\,\mathrm{Re}\left\{\mathrm{Tr}\left[H_{\mathrm{i,i+1}}G^<_{\mathrm{i+1,i}}(E)\right]\right\} \tag{8.9}$$

with $H_{\mathrm{i,i+1}}$ and $G^<_{\mathrm{i+1,i}}$ off-diagonal elements of the Hamiltonian matrix and lesser Green's function matrix.

8.2 Scattering models

In the previous section, we did not yet define the electron-phonon self-energies in Eqs. (8.1) and (8.4). In this section, we therefore present the expressions of the self-energies for the two types of electron-phonon interaction included in NEMO5: acoustic phonon scattering (Section 8.2.1) and non-polar optical phonon scattering (Section 8.2.2) [193]. We do not present their full derivation, but do mention the approximations made. We also discuss the lack of polar optical phonon scattering and suggest a way to mimic its effect (Section 8.2.3).

8.2.1 Acoustic phonons

The first kind of phonons included in NEMO5 are acoustic phonons. In our simulations, they are considered to be elastic, which means they do not change the energy of the charge carriers. Their dispersion relation is approximated as linear, with the slope determined by the sound velocity v_s (see Fig. 8.2). This is a reasonable approximation: because phonons are bosons, the occupation above an energy of $k_\mathrm{B}T/\hbar$ is negligible. This energy corresponds to the Debye wavelength:

$$q_\mathrm{D} = \frac{k_\mathrm{B}T}{\hbar v_\mathrm{s}} \tag{8.10}$$

As shown in Fig. 8.2, at low wavelengths, the acoustic phonon dispersion is almost linear. Besides the elasticity, an additional approximation is that only longitudinal modes are considered, which assumes a perfectly isotropic band at the Γ-point.

With these approximations, the retarded, lesser and greater self-energies for the interaction with the acoustic phonons at an energy E are the following [193]:

$$\Sigma^{R,<,>}_\mathrm{ac}(\alpha,\beta,E) = \frac{D^2_\mathrm{ac}k_\mathrm{B}T}{\rho v^2_\mathrm{s}}\delta_{\alpha,\beta}G^{R,<,>}(\alpha,\beta,E) \tag{8.11}$$

with ρ the density of the semiconductor and D_ac a deformation potential constant, which corresponds to a_c in the strain calculations of Chapter 6. Notice that the self-energies rely on their corresponding Green's function $G^{R,<,>}$, which themselves rely on the self-energies through Eqs. (8.1) and (8.4). NEMO5 therefore carries out a self-consistent loop between the calculation of the Green's functions and the self-energies, with the ballistic Green's functions as initial values. This self-consistent approach for calculating the electron-phonon scattering self-energies is known as the self-consistent Born method [56].

8.2.2 Non-polar optical phonons

The second kind of included phonons are the non-polar optical (NPO) phonons. They are considered to be inelastic and their dispersion is approximated by a constant phonon frequency ω_0 (see Fig. 8.2). As with the acoustic phonons, only longitudinal modes are considered. The corresponding retarded self-energy

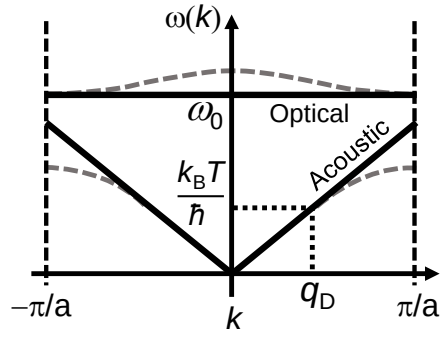

Figure 8.2: Schematic phonon dispersion, illustrating the approximations made for the acoustic and optical phonons. The dashed grey lines indicate a typical shape for an actual dispersion.

for emission and absorption of a phonon then reads [193]:

$$\Sigma^R_{\text{opt}}(\alpha,\beta,E) = \frac{\hbar D^2_{\text{opt}}}{2\rho\omega_0}\delta_{\alpha,\beta}\Big[(1+n_0)G^R(\alpha,\beta,E-\hbar\omega_0) + n_0 G^R(\alpha,\beta,E+\hbar\omega_0) + \frac{1}{2}G^<(\alpha,\beta,E-\hbar\omega_0) - \frac{1}{2}G^<(\alpha,\beta,E+\hbar\omega_0) + iP\int\frac{dE'}{2\pi}\left[\frac{G^<(\alpha,\beta,E-E')}{E'-\hbar\omega_0} - \frac{G^<(\alpha,\beta,E-E')}{E'+\hbar\omega_0}\right]\Big] \tag{8.12}$$

with D_{opt} the optical coupling constant, P the Cauchy principal value and $n_0 = 1/(e^{\hbar\omega_0/k_BT}-1)$ the Bose-Einstein distribution of the phonons. The corresponding lesser and greater self-energies are [193]:

$$\Sigma^{<,>}_{\text{opt}}(\alpha,\beta,E) = \frac{\hbar D^2_{\text{opt}}}{2\rho\omega_0}\delta_{\alpha,\beta}\Big[n_0 G^{<,>}(\alpha,\beta,E-\hbar\omega_0) + (1+n_0)G^{<,>}(\alpha,\beta,E+\hbar\omega_0)\Big] \tag{8.13}$$

As for the acoustic phonons, these self-energies also depend on the Green's functions and are obtained with the self-consistent Born method.

8.2.3 Polar optical phonons

NEMO5 lacks the ability to model scattering by polar optical phonons (POP). This type of scattering occurs in polar materials; the relative movement of the

lattice atoms creates a local electric field which interacts with the charge carriers [194]. Although its exact strength is unknown, POP scattering is expected to be strong in III-V materials with polar bonds, such as the materials we are considering. Including only NPO scattering in the simulations is therefore likely an underestimation of the actual electron-phonon scattering strength in these materials. Unfortunately, POP scattering is difficult to model, because it is a long range non-local interaction. The corresponding self-energy is non-diagonal in position, which reduces the sparsity of the NEGF system and prevents the straightforward use of RGF to speed up the inversion.

To compensate for the lack of POP scattering in NEMO5, we use D_{opt} in the NPO self-energy as an empirical parameter to increase the scattering strength. In contrast with the deformation potential constant D_{ac} for the acoustic phonons, D_{opt} is not well known for III-V materials. Theoretically, D_{opt} is even zero because of symmetry if the electrons are located in an isotropic valley at the Γ-point [194]. This is the case in most III-V materials. In an actual device, however, this symmetry is broken because of confinement. The consequence is that D_{opt} cannot be calculated from the simplified theory for III-V materials and therefore no values are listed in literature. We therefore start from the value for Si and gradually increase the scattering strength, to assess the sensitivity of the investigated configurations to the scattering. This approach is approximative since it neglects the non-local nature of the POP scattering.

8.3 Simulation results

Having discussed the simulation procedure and scattering models employed by NEMO5, we now compare the resonant TFET with a conventional heterostructure TFET. We first compare the configurations with ballistic simulations, and then add electron-phonon scattering with increasing strength to assess the difference in sensitivity of the transfer characteristics.

8.3.1 Simulation details

The simulated configurations of the conventional heterostructure TFET and the resonant TFET are depicted in Fig. 8.3, with the details in Table 8.1. The conventional TFET consists of a GaSb/InAs heterostructure. This material combination has a broken band alignment in bulk, but for the simulated T_{body} of 1.8 nm, the alignment is staggered as a result of SIQC-induced bandgap widening. The resonant TFET structure was designed by Long *et al.* [66] and consists of several material layers, where the tunnel junction is located at the

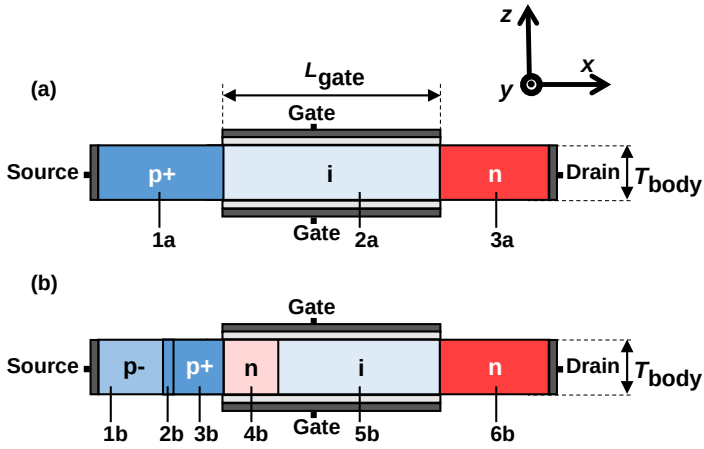

Figure 8.3: Investigated configurations: (a) conventional heterostructure pointTFET and (b) resonant TFET.

Table 8.1: Details of the simulated configurations (see Fig. 8.3 for the region numbers).

Region - Material	Length [nm]	Doping [cm^{-3}]
1a - GaSb	9	$5x10^{19}$
2a - InAs	30.5	$1x10^{15}$
3a - InAs	17.3	$5x10^{19}$
1b - AlSb	4.6	$3x10^{19}$
2b - $AlGa_{0.5}Sb_{0.5}$	1.2	$6x10^{19}$
3b - GaSb	3.2	$5x10^{19}$
4b - InAs	3.4	$1x10^{15}$
5b - AlInAsSb	27.1	$1x10^{15}$
6b - AlInAsSb	17.3	$5x10^{19}$
Channel orientation	[110]	
L_{gate} [nm]	30.5	
T_{body} [nm]	1.8	
EOT [nm]	0.8	
GWF [eV]	4.57	
Phonon energy ($\hbar\omega_0$) [eV]	0.030	
ρ [kg/m^3]	2336	
v_s [m/s]	8344	
D_{ac} [eV]	9	
D_{opt} [eV/nm]	varying	

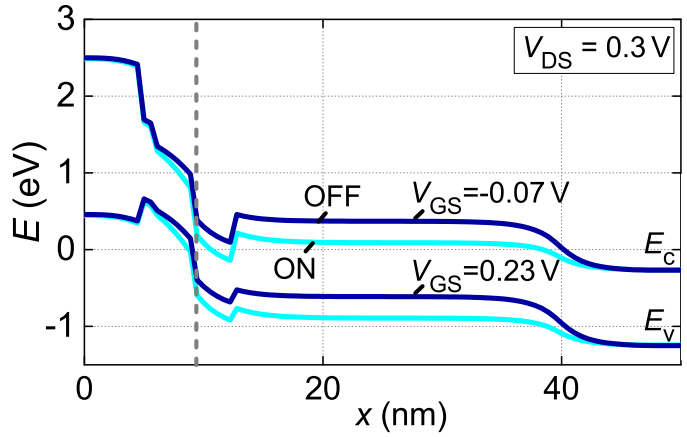

Figure 8.4: Band diagram of a resonant TFET in the OFF and ON-state. The tunnel junction is indicated with a dashed line.

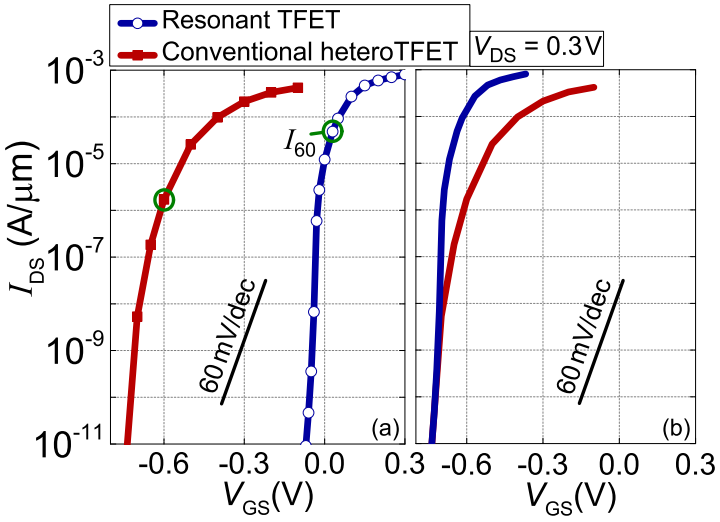

Figure 8.5: Ballistic QM simulated transfer characteristics, comparing a conventional GaSb/InAs pointTFET with a resonant TFET. More configuration details in Fig. 8.3 and Table 8.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $1\mathrm{x}10^{-11}\mathrm{A}/\mu\mathrm{m}$ coincides.

GaSb/InAs interface. In the source and channel, two wide bandgap materials (AlSb and AlInAsSb) provide a band offset that creates two quantum wells at the tunnel junction. The device turns to the ON-state when the quantized states in these wells overlap and resonant tunneling becomes possible (see Fig. 8.4). A thin transition region has also been included between the AlSb and the GaSb.

The direction of transport in the simulated configurations is oriented along the [110]-direction, with confinement in the $(1\bar{1}0)$-plane, unlike the rest of this thesis where transport is in the [100]-direction with confinement in the (001)-plane. This choice offers a higher transmission probability in strongly confined structures [66]. Because of the confinement in the $(1\bar{1}0)$-plane, the resulting hole effective mass is smaller than for the (001)-confinement, which improves the transmission probability. Additionally, the tunnel barrier is slightly thinner. This is because the heavy hole effective mass is larger in the [110]-direction than in the [100]-direction, which results in a smaller shift of the valence band edge under $(1\bar{1}0)$-confinement.

The simulations of these configurations with electron-phonon scattering are computationally very demanding for several reasons. First, as discussed in Section 8.1, there are two self-consistent loops which each have to converge. Second, to capture the transmission resonances in the resonant TFET, a fine energy mesh is required. NEMO5 adaptively refines the mesh at each Poisson iteration to ensure the resonances are resolved properly. Third, the simulation time of a single iteration increases rapidly with the number of atoms in the simulated structure, since each mesh point corresponds to an atom position. To limit the computational time, T_{body} is therefore chosen very confined at only 1.8 nm, which is three crystal unit cells. Even so, the scattering simulations still require a large number of cores on a supercomputer, around 2000, to keep the simulation time of a single bias point within a period of hours. Thanks to the strongly parallelized structure of NEMO5, the calculation time scales almost linearly with the number of cores. The simulations were carried out on the Stampede supercomputer at the Texas Advanced Computing Center of the University of Texas at Austin, TX, USA.

8.3.2 Ballistic conventional TFET versus resonant TFET

As a starting point, we compare the resonant TFET to the conventional TFET with ballistic simulations. Fig. 8.5 shows a much lower SS for the resonant TFET than the conventional TFET, in line with results in literature [66]. The steep onset is the result of resonant tunneling in the quantum wells around the tunnel junction. When the energy levels of the first confined state in these wells overlap, the transmission probability rises sharply and becomes close to

one. The onset of the conventional TFET is more gradual, as it depends on the gradual thinning of the tunnel barrier.

The resonant tunneling hence alleviates the reduction in transmission probability due to the SIQC-induced bandgap widening. In the conventional TFET, SIQC turns the band alignment from a broken alignment, which has a high transmission probability, to a staggered alignment with a significant effective bandgap and thus a lower transmission probability. With the resonant tunneling, the transmission probability is close to one when the resonant states overlap, regardless of SIQC-induced increase in the tunnel barrier at the junction.

8.3.3 Electron-phonon scattering in conventional TFET versus resonant TFET

When we introduce NPO electron-phonon scattering into the resonant TFET simulations, the SS degrades noticeably with increasing scattering strength as shown in Fig. 8.6, while the acoustic phonons only result in a shift of the transfer characteristics. D_{opt} is increased in steps with a factor of $\sqrt{2}$ from the value for Si (110 eV/nm)[195] up to the point at which the self-consistent loop no longer converges (220 eV/nm). The factor $\sqrt{2}$ corresponds to a factor of 2 increase in scattering strength (see the prefactor of the NPO self-energy in Eq. (8.12)). The SS degradation worsens as D_{opt} is increased, although the SS remains sub-60 mV/dec.

The origin of the NPO-induced SS degradation is explained in Fig. 8.7, which shows a blurring of the band edges in the local DOS (LDOS) around the tunnel junction. This blurring is a consequence of the inelastic nature of the NPO scattering, which allows electrons to achieve energies above the valence band edge. Additionally, Fig. 8.7 shows that one of the quantized states in the second quantum well has a large tail into the source bandgap. At onset, instead of an abrupt overlap of this quantized state with the valence band edge, the quantized state gradually overlaps with the blurred valence band edge. During the initial stages of this overlap, electrons from the valence band in the source are able to pass into the channel with the energetic aid of a phonon, before the band bending is sufficient for ballistic tunneling. These phonon-assisted transitions result in a tail to the transfer characteristics, thereby degrading the SS.

In contrast to the resonant TFET, the SS of the conventional TFET is much less affected by the NPO scattering, as shown in Fig. 8.8. Although the transfer characteristics have shifted due to the acoustic phonon scattering, the SS is largely unaffected. This is to be expected, since in the conventional TFET,

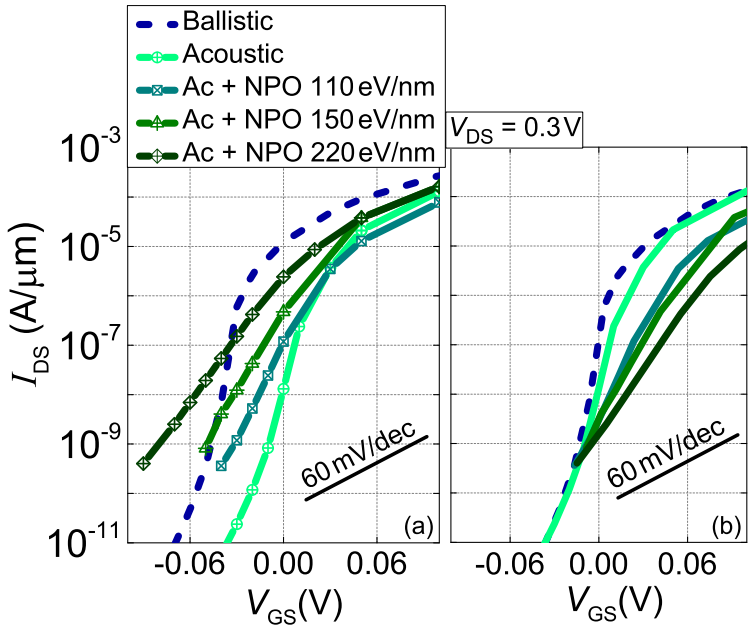

Figure 8.6: QM simulated transfer characteristics of a resonant TFET with varying strength of NPO electron-phonon scattering. More configuration details in Fig. 8.3(b) and Table 8.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is $4x10^{-10}$A/μm coincides.

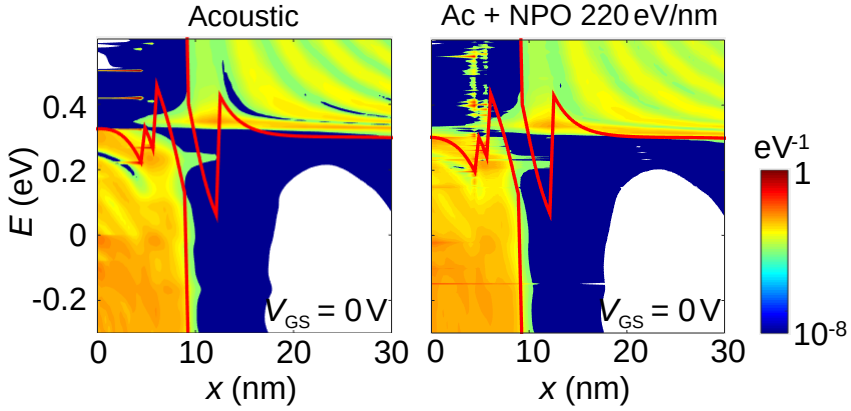

Figure 8.7: LDOS around the tunnel junction of a resonant TFET with (a) acoustic phonon scattering and (b) acoustic and NPO phonon scattering. The band diagram is superimposed in solid lines.

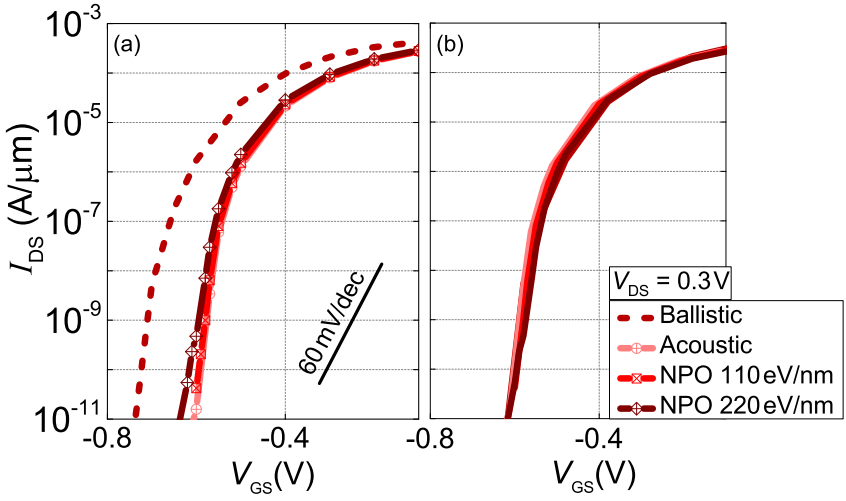

Figure 8.8: QM simulated transfer characteristics of a conventional GaSb/InAs pointTFET with varying strength of NPO electron-phonon scattering. More configuration details in Fig. 8.3(a) and Table 8.1. (a) Unshifted and (b) shifted characteristics such that the V_{GS} at which I_{OFF} is 1x10^{-11}A/μm coincides.

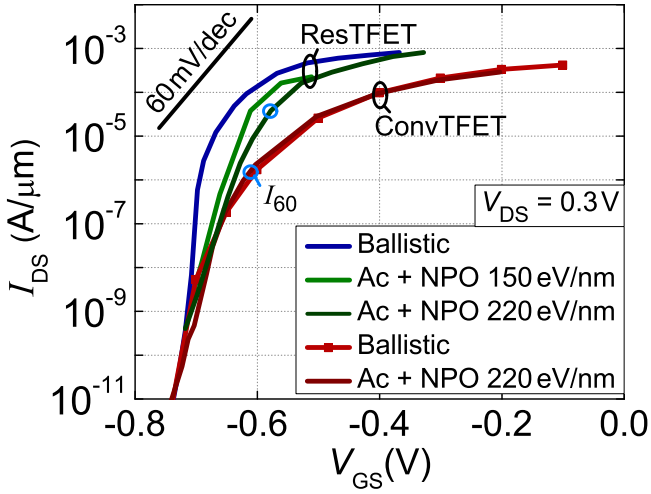

Figure 8.9: QM simulated transfer characteristics, comparing a conventional GaSb/InAs pointTFET (convTFET) with a resonant TFET (resTFET) for varying scattering strength. More configuration details in Fig. 8.3 and Table 8.1. The curves are shifted such that V_{GS} at an I_{OFF} of 4x10^{-10}A/μm coincides.

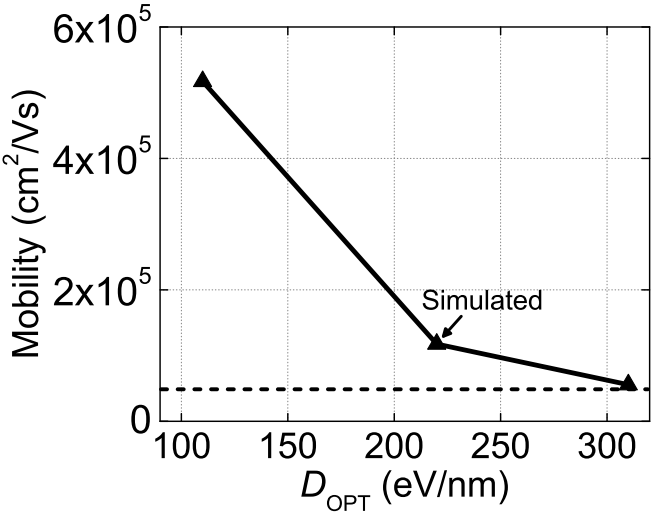

Figure 8.10: Mobility estimations corresponding to different values of D_{opt}. The dashed values indicate the experimental bulk mobility of InAs.

there is no quantized state which can leak into the source bandgap as in the resonant TFET and provide phonon-assisted leakage paths.

Although the conventional TFET is less sensitive to the NPO scattering, Fig. 8.9 shows that the performance of the resonant TFET is still superior even for the largest scattering strength. For a supply voltage window of 0.3 V, the resonant TFET obtains an I_{ON} of around 370 μA/μm with an I_{60} of 40 μA/μm, compared to 50 μA/μm and 2 μA/μm respectively for the conventional TFET.

8.4 Mobility estimation

In the previous section it became clear that the performance of the resonant TFET depends strongly on the chosen scattering strength, which we have varied somewhat arbitrarily through the empirical parameter D_{opt}. To get a feeling for a realistic magnitude of D_{opt}, we now estimate to which bulk mobilities the simulated values of D_{opt} correspond and compare them to the experimentally measured bulk mobility.

The bulk mobilities are extracted from simulations of quasi-1D lowly doped

($1x10^{15}cm^{-3}$) InAs samples with a varying length L_S. The samples are considered to be infinite in the two orthogonal directions, and are assumed to exhibit rotational symmetry in the orthogonal k-space. The mobility can then be calculated as [196]:

$$\mu = \frac{L_S}{n_{1D} q R_S} \tag{8.14}$$

with n_{1D} the averaged 1D electron density (in units of m^{-1}) and R_S the sample resistance. The latter can be separated from the contact area resistance R_c and the ballistic resistance R_0 by simulating samples of length $N_S \cdot L_S$ with N_S a scaling factor. While R_0 and R_c remain constant with L_S, the total resistance changes as a linear function of N_S:

$$R(N_S) = R_0 + R_c + N_S R_S \tag{8.15}$$

Simulating the resistance for at least two sample lengths therefore allows the extraction of R_S.

Fig. 8.10 shows the mobilities corresponding to the values of D_{opt} in Figs. 8.6 and 8.8. The largest simulated value of D_{opt} for which the self-consistent loop converges, corresponds to a mobility of $1.1x10^5$ cm^2/Vs, which is about two times higher than the experimental bulk mobility value of around $4x10^4$ cm^2/Vs [178]. Note that our calculated mobility only takes into account electron-phonon scattering, while the experimental value includes other scattering mechanisms as well, for instance due to surface roughness and impurities. We can therefore conclude that our simulations probably slightly underestimate the scattering strength.

8.5 Conclusion

In this chapter, we compared the sensitivity to electron-phonon scattering of a resonant TFET to a conventional heterostructure TFET with the NEMO5 package. We discussed how NEMO5 includes elastic acoustic and inelastic NPO phonon scattering through self-consistent self-energies in the NEGF system. We noted that POP scattering is not included due to the excessive computational demand, although it is expected to be important in polar III-V materials. We therefore used the optical coupling constant as an empirical parameter to probe higher scattering strengths and thereby mimic the effect of POP scattering.

We first verified the superior ballistic performance of the resonant TFET compared to the conventional TFET, which is thanks to the abrupt overlap of resonant states in quantum wells near the tunnel junction. Next, we showed that the elastic acoustic phonon scattering results in a mere shift of the transfer

characteristics of both configurations, while the NPO scattering degrades the SS especially of the resonant TFET. The origin of this larger scattering sensitivity of the resonant TFET was found to be inelastic phonon-assisted transitions between the valence band in the source and a tail of the channel quantized state that penetrates the source bandgap. Even so, the resonant TFET still outperformed the conventional TFET for the largest scattering strength.

Finally, we linked the values of the optical coupling constant to a bulk mobility to get a feeling for the simulated scattering strengths. The highest scattering strength for which the simulations converged was found to correspond to a mobility about two times higher than the experimental bulk value. This seems to indicate that the presented results still somewhat underestimate the scattering.

In light of the findings of this chapter, the resonant TFET remains an interesting configuration, although it is complicated to experimentally fabricate due to the many heterojunctions. Future research should focus on reducing the impact of inelastic scattering by limiting the penetration of the channel resonant state in the source bandgap, e.g. by choosing a source material with a large attenuation.

Chapter 9

Conclusions and outlook

This chapter summarizes the most important conclusions of this thesis (Section 9.1) and provides an outlook with suggestions for future research (Section 9.2).

9.1 Conclusions

The summary and conclusions are organized with their corresponding key goal as formulated in the introduction in Chapter 1. The first set of conclusions therefore focuses on the formalism, while the second and third set contain design optimization conclusions.

1. **Develop a QM formalism to simulate transport in heterostructure TFETs in a computationally efficient way.**

   In Chapter 3, we combined a fully QM ballistic multi-band EF band structure model with the QTBM to describe BTBT, the TFET's active current component, for 2D potentials. We constructed an EF system, and solved it with a WF approach. The EF system incorporates a previously developed heterostructure transformation of the basis functions. Numerically, we discretized the EF system with a finite difference scheme.

   In Chapter 4, we implemented our formalism for a two-band band structure model. We highlighted the lack of coupling between the bands in the directions orthogonal to transport, which results in an erroneous upward

curvature of the valence bands in those directions. We reintroduced the correct curvature by adding an effective mass term to the EF system. We found that the model still contains out-of-zone spurious solutions, which we decided not to inject, but keep in the boundary conditions for continuity of the EFs. The parameters of the two-band model were calibrated with experimental $In_{0.53}Ga_{0.47}As$ p-i-n diodes. The calibration showed that both the QM and SC models could reproduce the BTBT current in wide diodes.

In Chapter 5, we expanded the implementation to a full-zone fifteen-band model. We replaced the finite difference scheme in the confined direction with a spectral approach, which reduces the computational burden and allows to remove spurious solutions near the zone-center. We also developed a parameter fitting procedure for multi-band models that retains the commutativity of the momentum matrices. This enabled computationally efficient full-zone simulations of homo- and heterostructure TFETs. We compared fifteen-band results of a homostructure $In_{0.53}Ga_{0.47}As$ p-i-n TFET to experiment and found a better match than SC simulations.

In Chapter 6, we added spin and mechanical strain to the band structure description, resulting in a thirty-band model. Strain was incorporated with a position-dependent eight-band strain Hamiltonian, approximated for low values of k. This enabled the incorporation of an arbitrary strain tensor, to simulate both uniform or non-uniform strain profiles across the device structure.

In Chapter 7, we coupled the solutions of the fifteen-band EF system to Poisson's equation in a self-consistent loop. We applied a combination of a delta-normalization in the open directions and a box normalization in the confined direction. We implemented a Gummel scheme with adaptive damping to achieve convergence between carrier density and electrostatic potential.

All of this functionality was bundled in a QM simulator called Pharos, after the lighthouse of the city of Alexandria, which was one of the most important centers of knowledge in classical antiquity thanks to its extensive library. Pharos as it stands today allows for full-zone simulation and optimization of a wide variety of TFET configurations: homostructure and heterostructure, very wide (50 nm) to very confined (5 nm), lineTFET and pointTFET, relaxed and (non-uniformly) strained.

Next to transfer characteristics, which have been the focus of this thesis, also output characteristics are being investigated in the imec TFET team with Pharos. As discussed in Section 2.3.3, the output characteristics should demonstrate saturation with a saturation voltage as low as possible, and a linear onset. Combined with charge densities and capacitances, which can also be extracted from the developed simulator, this enables simulations of simple TFET circuits, which allows for an energy-based comparison with MOSFET circuits.

2. **Assess the viability of direct bandgap III-V TFET.**

   With the two-band implementation of Chapter 4, we carried out a preliminary study of III-V staggered bandgap heterostructure pocketed pointTFETs for different values of the body thickness and found a counteracting effect between SIQC and gate control. As the body thickness was decreased to very confined dimensions, below 10 nm, we found that the TFET performance suffered. This suggests an optimum exists for the body thickness.

   With the extended fifteen-band model of Chapter 5, III-V pocketed lineTFET and pointTFET could be compared. We first verified that the lineTFET concept holds up in direct bandgap III-V materials by comparing QM to SC simulations. We found that the FIQC in the QM case only induced a shift of the transfer characteristics as compared to the SC results, with no degradation in the SS. Comparing optimized n-channel pocketed lineTFETs and pointTFETs, we found a smaller difference in performance between the two configurations than earlier studies for group IV materials, with the pointTFET matching the lineTFET performance. For highly optimized configurations, the type of configuration did not make a large difference. The higher fabrication complexity of the lineTFET therefore gives the advantage to the pointTFET as the preferred III-V configuration.

   Both lineTFET and pointTFET configurations performed poorly as a pTFET, none of them reaching sub-60 mV/dec SS due to large source doping degeneracy. We therefore proposed an improved source design, which consists of creating a source region with low degeneracy, either with a low doping or a heterostructure, while retaining the high electric field at the tunnel junction with a dopant pocket. P-channel pointTFET configurations with the improved source were able to match and even surpass n-channel pointTFET performance, enabling complementary circuit implementations. This inventive configuration had a strong

positive impact on the III-V TFET community, as the perspective of a pMOSFET or an extremely complex pTFET design were not very appealing alternatives.

Heterostructure nTFETs and pTFETs, the latter with improved source design, were optimized by imec team members with Pharos and showed promising results for the staggered bandgap $GaAs_{0.5}Sb_{0.5}/In_{0.53}Ga_{0.47}As$ system. Values of I_{ON} above 10^2 $\mu A/\mu m$ with I_{60} reaching around 10 $\mu A/\mu m$ for a V_{DD} of 0.3 V with I_{OFF} below $1x10^{-11}$ $A/\mu m$ for the nTFETs and below $1x10^{-10}$ $A/\mu m$ for the pTFETs are promising to compete with MOSFET. These device simulations cleared the way for the next step towards a TFET technology, namely predictive circuit simulations.

In Chapter 8, we used NEMO5 to compare the impact of electron-phonon scattering on a resonant TFET to a conventional heterostructure TFET. We found non-polar optical phonon scattering to degrade the SS of the resonant TFET, while the conventional TFET was almost unaffected. The resonant TFET did, however, show the lowest SS even for the highest scattering strength.

3. **Investigate whether strain can be a performance booster for III-V TFETs.**

With the strain implementation of the thirty-band model in Chapter 6, we confirmed reports in literature that the SS of uniformly strained lattice-matched heterostructure TFETs degrades, both for biaxial and uniaxial stress. The origin of the degradation was found to be a strain-induced reduction in valence band DOS in the source, resulting in an increased source degeneracy. With the improved source design of Chapter 5, we were able to counter the SS degradation, which enabled us to take full benefit of the improvement in I_{ON} for the biaxially stressed configurations.

For heterostructure TFETs with a non-uniform strain profile due to lattice mismatch, we found that a proper selection of the material combination leads to a smaller $E_{G,eff}$ and hence an improved I_{ON} over the lattice-matched reference. We confirmed that it is indeed the reduction in $E_{G,eff}$ and not the non-uniformity of the strain profile which causes the improvement. We found no degradation of the SS as in the uniform case, because if channel and drain are grown pseudomorphically on a relaxed source, the non-uniform strain profile in the source remains limited to a region close to the tunnel junction. The source degeneracy is thereby

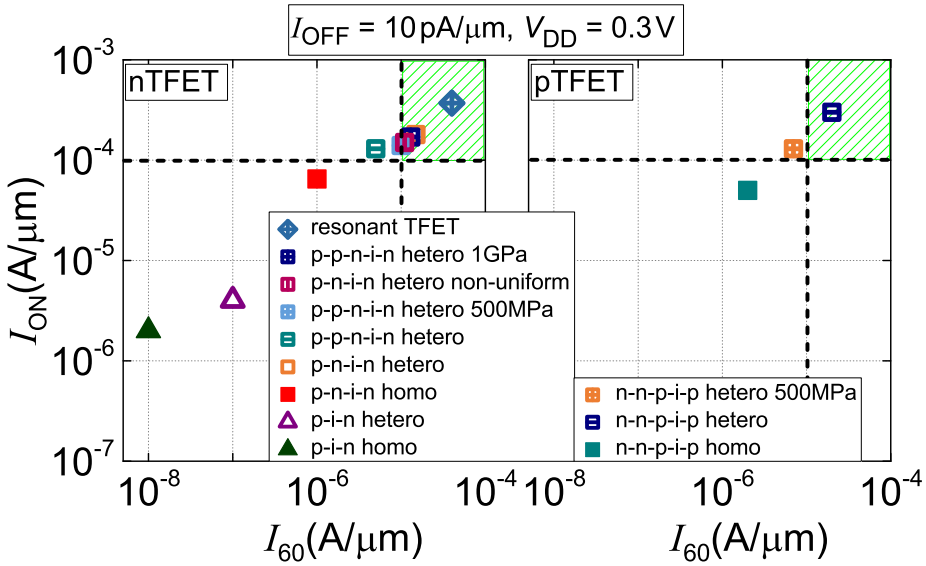

Figure 9.1: Overview of performance of representative TFET configurations investigated in this thesis. The dashed lines indicate the minimum target specifications for transfer characteristics defined in Section 2.6.

unaffected. We therefore identified lattice mismatch as an additional tuning parameter, expanding the heterostructure TFET design space.

In Fig. 9.1, we summarize the most representative simulated pointTFET transfer characteristics in this thesis, by plotting I_{ON} versus I_{60}. The main goal of this figure is to show the trend of improving performance, starting from a simple homostructure p-i-n structure and moving to heterostructures, pockets and strained configurations. Only the most optimized, but also most experimentally complicated configurations reach the minimum target specifications outlined in Section 2.6. Notice that for the pTFET, the improved source design is absolutely necessary to reach the minimum specs, since conventional source configurations do not even make it to the plot because of a lack of sub-60 mV/dec SS.

9.2 Outlook

9.2.1 Suggestions for future research

Confined nanowire implementations of TFETs with the gate wrapped around the channel seem promising because of the good electrostatic control on the tunnel junction. To simulate these structures, the translationally invariant direction in the present implementation of Pharos should be replaced with a second confined direction. Our spectral approach could then be applied in this direction analogous to the other confined direction to limit the computational demand and avoid spurious solutions. Because of the strong confinement, self-consistency would also be indispensable, and it might be required to take into account the penetration of the wave function into the gate oxide. The latter could potentially be introduced as an artificial wide bandgap semiconductor.

More challenging is the QM modeling of non-idealities and parasitic current mechanisms. Most notably TAT and dopant tails form a threat to the TFET SS, and are at this point neglected in most QM simulation studies. These phenomena are challenging to model, because they originate from localized states, formed by defects in the material. A first step could be to include traps as local potential variations in a ballistic treatment. In a next step, electron-phonon coupling could be introduced perturbatively to model the thermal leakage currents through the trap levels.

9.2.2 General outlook

At the end of this PhD project in 2016, TFET research is again at a crossroads, albeit not the same as at the beginning. In this thesis and other publications, promising III-V TFET configurations have been identified based on their ideal transfer characteristics. The question is now: will experimental TFETs be able to overcome non-idealities and actually demonstrate performance close to these ideal predictions, both for the transfer and output characteristics? The understanding gained from accurate QM modeling of the TFET non-idealities will be key in finding ways to reduce parasitic effects, close the gap between simulations and experiments, and enable the TFET to fulfill its original purpose as a low-voltage MOSFET successor.

Appendix A

Matrix construction

This appendix provides instructions on how to construct the discretized matrices for the solution of the EF system and Poisson's equation. We first present the discretized form of the general EF system matrix $\mathbf{K}$ of Eq. (3.49) (Section A.1). Then we show the modifications to $\mathbf{K}$ required for our two-band model, which introduces effective masses in the orthogonal directions (Section A.2). Next, we show the EF system matrix with a spectral approach in the confined direction, which is used for higher band models and in the presence of strain (Section A.3). We also include the momentum matrices (Section A.4) and spin-orbit matrix (Section A.5), necessary for the construction of the fifteen- and thirty-band system matrices. Finally, we discuss the discretization and solution of Poisson's equation in the Gummel method of Eq. (7.26) (Section A.6).

The construction of the various matrices relies heavily on the Kronecker product, denoted with the symbol $\otimes$.

A.1 General discretized EF system

We first present the discretized form of the general EF system matrix in Eq. (3.15), which was derived in Chapter 3. The second order derivative matrices can be written as:

$$\mathbf{D2}_{x,\mathrm{tot}} = \mathbf{D2}_x \otimes \mathbf{I}_z \otimes \mathbf{I}_\mathrm{N} \tag{A.1}$$

$$\mathbf{D2}_{z,\mathrm{tot}} = \mathbf{I}_x \otimes \mathbf{D2}_z \otimes \mathbf{I}_\mathrm{N} \tag{A.2}$$

with $\mathbf{D2}_x$ defined in Eq. (3.35) and $\mathbf{D2}_z$ analogous. $\mathbf{I}_x$, $\mathbf{I}_z$ and $\mathbf{I}_\mathrm{N}$ are the $N_x \times N_x$, $N_z \times N_z$ and $N \times N$ unity matrices respectively.

For the first order derivative matrices, we have:

$$\mathbf{D1}_{x,\mathrm{tot}} = \mathbf{D1}_x \otimes \mathbf{I}_z \otimes \mathbf{I}_\mathrm{N} \tag{A.3}$$

$$\mathbf{D1}_{z,\mathrm{tot}} = \mathbf{I}_x \otimes \mathbf{D1}_z \otimes \mathbf{I}_\mathrm{N} \tag{A.4}$$

with $\mathbf{D1}_x$ defined in Eq. (3.33) and $\mathbf{D1}_z$ analogous. For the interband momentum matrix terms, we can then construct:

$$\mathbf{P}_{x,\mathrm{tot}} = \mathbf{D1}_x \otimes \mathbf{I}_z \otimes \mathbf{P}_x \tag{A.5}$$

$$\mathbf{P}_{y,\mathrm{tot}} = \mathbf{I}_x \otimes \mathbf{I}_z \otimes \mathbf{P}_y \tag{A.6}$$

$$\mathbf{P}_{z,\mathrm{tot}} = \mathbf{I}_x \otimes \mathbf{D1}_z \otimes \mathbf{P}_z \tag{A.7}$$

with $\mathbf{P}_x$, $\mathbf{P}_y$ and $\mathbf{P}_z$ the interband momentum matrices.

The two remaining terms are the bulk Hamiltonian matrix term:

$$\mathbf{H}_\mathrm{tot} = \mathbf{I}_x \otimes \mathbf{I}_z \otimes \mathbf{H} \tag{A.8}$$

with $\mathbf{H}$ the bulk Hamiltonian matrix, and the external potential energy term:

$$\mathbf{V}_{e,\mathrm{tot}} = \mathrm{diag}(\mathbf{V}_e) \otimes \mathbf{I}_\mathrm{N} \tag{A.9}$$

with $\mathbf{V}_e$ an $N_x \cdot N_z$ vector of the discretized external potential energy.

Finally, we combine these matrices to construct the full discretized EF system matrix:

$$\begin{aligned}\mathbf{K} = \frac{-\hbar^2}{2m_e}\mathbf{D2}_{x,\mathrm{tot}} + k_y^2\frac{\hbar^2}{2m_e}\mathbf{I}_\mathrm{tot} - \frac{\hbar^2}{2m_e}\mathbf{D2}_{z,\mathrm{tot}} - \frac{i\hbar}{m_e}\mathbf{P}_{x,\mathrm{tot}} + k_y\frac{\hbar}{m_e}\mathbf{P}_{y,\mathrm{tot}} \\ - \frac{i\hbar}{m_e}\mathbf{P}_{z,\mathrm{tot}} + \mathbf{H}_\mathrm{tot} + \mathbf{V}_{e,\mathrm{tot}}\end{aligned} \tag{A.10}$$

with $\mathbf{I}_\mathrm{tot}$ the $N_x \cdot N_z \cdot N \times N_x \cdot N_z \cdot N$ unity matrix.

A.2 Two-band discretized EF system

In the two-band model discussed in Chapter 4, the lack of coupling in the orthogonal directions requires the introduction of effective masses into the EF

system (see Eq. (4.9)). To introduce these in the discretized system, we define the following diagonal matrix:

$$\mathbf{M} = \mathrm{diag}\left(\mathbf{M}_{1,1}\mathbf{M}_{1,2}\cdots\mathbf{M}_{N_x,N_z}\right) \tag{A.11}$$

with

$$\mathbf{M}_{i,j} = \left[\frac{1}{m_{\mathrm{c}}^*(x_i,z_j)}\,\frac{1}{m_{\mathrm{v}}^*(x_i,z_j)}\right] \tag{A.12}$$

The discretized EF system matrix of Eq. (A.10) is then modified to:

$$\mathbf{K} = \frac{-\hbar^2}{2m_e}\mathbf{D2}_{x,\mathrm{tot}} + k_y^2\frac{\hbar^2}{2m_e}\mathbf{M} - \frac{\hbar^2}{2m_e}\mathbf{M}.\mathbf{D2}_{z,\mathrm{tot}} - \frac{i\hbar}{m_e}\mathbf{P}_{x,\mathrm{tot}} + \mathbf{H}_{\mathrm{tot}} + \mathbf{V}_{e,\mathrm{tot}}. \tag{A.13}$$

A.3 Spectral discretized EF system

Starting with the fifteen-band model discussed in Chapter 5, the finite difference discretization in the confined z-direction is replaced with a spectral decomposition (see Eq. (5.8)). To construct the discretized spectral EF system, the following matrices are redefined:

$$\mathbf{D2}_{x,\mathrm{tot}} = \mathbf{D2}_x \otimes \mathbf{I}_{k_z} \otimes \mathbf{I}_{\mathrm{N}}$$

$$\mathbf{P}_{x,\mathrm{tot}} = \mathbf{D1}_x \otimes \mathbf{I}_{k_z} \otimes \mathbf{P}_x \tag{A.14}$$

$$\mathbf{P}_{y,\mathrm{tot}} = \mathbf{I}_x \otimes \mathbf{I}_{k_z} \otimes \mathbf{P}_y \tag{A.15}$$

with N_{k_z} the number of spectral components and $\mathbf{I}_{k_z}$ the $N_{k_z} \times N_{k_z}$ unity matrix. We additionally define a diagonal matrix containing the values of k_z:

$$\mathbf{KZ} = \mathbf{I}_x \otimes \mathrm{diag}\left(k_{z1}, k_{z2}\cdots k_{zN_{kz}}\right) \tag{A.16}$$

For the external potential energy term, we denote $\mathbf{V}_i$ as the vector containing the N_z values of $\mathbf{V}_e$ for the x-value x_i. $\tilde{\mathbf{V}}_i$ is then the corresponding transformed vector in the spectral domain, containing N_{k_z} spectral components. With these definitions, we write the external potential energy term as:

$$\mathbf{V}_{e,\mathrm{tot}} = \sum_i^{N_x}\left[\mathbf{J}_{i,x} \otimes \hat{P}_{k_z}\left(\hat{T}\tilde{\mathbf{V}}_{i,1:N_{kz}-2} \mp \hat{T}\tilde{\mathbf{V}}_{i,N_{kz}:3}\mathbf{A}_{k_z}\right) \otimes \mathbf{I}_{\mathrm{N}}\right] \tag{A.17}$$

with $\hat{T}$ an operator that turns a vector into a Toeplitz matrix, $\hat{P}_{k_z}$ an operator that pads a matrix with zeros at each side until it has dimensions $N_{k_z} \times N_{k_z}$, $\mathbf{A}_{k_z}$ the $N_{k_z} \times N_{k_z}$ anti-diagonal identity matrix and $\mathbf{J}_{i,x}$ an $N_x \times N_x$ matrix with one non-zero element: a 1 on the i'th position of the diagonal. The minus (plus) sign corresponds to the odd (even) EFs.

In the thirty-band model discussed in Chapter 6, a spin-orbit term and a strain term are added to the bulk Hamiltonian term of the EF system (see Eq. (6.10)). In matrix form, we write:

$$\mathbf{H}_{\text{cmb}} = \mathbf{H} + \mathbf{H}_{\text{SO}} + \mathbf{H}_{\text{S}} \tag{A.18}$$

with $\mathbf{H}_{\text{cmb}}$ an $N \times N$ matrix. $\mathbf{H}_{\text{cmb}}$ varies with position in both the x- and z-direction. Analogous to the potential energy term in the previous section, we therefore define a vector $\mathbf{G}_{i,nm}$, which contains the variation of the matrix element $[\mathbf{H}_{\text{cmb}}]_{nm}$ over the z-direction at x_i, and $\tilde{\mathbf{G}}_{i,nm}$, which contains the associated spectral components. We can then write the total bulk Hamiltonian term of the spectral EF system Eq. (6.11) as:

$$\mathbf{H}_{\text{tot}} = \sum_{i}^{N_x} \left[\mathbf{J}_{i,x} \otimes \sum_{n,m}^{N} \hat{P}_{k_z} \left(\hat{T}\tilde{\mathbf{G}}_{i,nm,1:N_{kz}-2} \mp \hat{T}\tilde{\mathbf{G}}_{i,nm,N_{kz}:3} \mathbf{A}_{k_z} \right) \otimes \mathbf{J}_{nm,\text{N}} \right] \tag{A.19}$$

where $\mathbf{J}_{nm,\text{N}}$ is an $N \times N$ matrix where the element nm is 1 and the other elements are zero. Eq. (A.19) can be inserted in the discretized EF system matrix of Eq. (A.20). We can then finally construct the resulting discretized spectral EF system matrix as:

$$\begin{aligned} \mathbf{K} = \frac{-\hbar^2}{2m_e}\mathbf{D2}_{x,\text{tot}} + k_y^2\frac{\hbar^2}{2m_e}\mathbf{I}_{\text{tot}} + \frac{\hbar^2}{2m_e}\left(\mathbf{KZ}^2 \otimes \mathbf{I}_{\text{N}}\right) - \frac{i\hbar}{m_e}\mathbf{P}_{x,\text{tot}} + k_y\frac{\hbar}{m_e}\mathbf{P}_{y,\text{tot}} \\ \pm \frac{i\hbar}{m_e}\left(\mathbf{KZ} \otimes \mathbf{P}_z\right) + \mathbf{H}_{\text{tot}} + \mathbf{V}_{e,\text{tot}} \end{aligned} \tag{A.20}$$

A.4 Momentum matrices

This section contains the momentum matrices used in the fifteen- and thirty-band models. The columns and rows correspond to the following basis states (in order):

$$|S_v\rangle, |Y\rangle, |Z\rangle, |X\rangle, |S\rangle, |Z_c\rangle, |Y_c\rangle, |X_c\rangle, |S_u\rangle, |D_z\rangle, |D_x\rangle, |Y_d\rangle, |Z_d\rangle, |X_d\rangle, |S_q\rangle \tag{A.21}$$

The matrix elements connecting these basis states are defined in Section 5.1. The fifteen-band momentum matrices are then defined as:

$$\mathbf{P}_x = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & P_{\mathrm{S}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -P & 0 & 0 & 0 & 0 & -P_3 & \sqrt{3}\,P_3 & 0 & 0 & 0 & -P_2 \\
0 & 0 & 0 & -P^* & 0 & 0 & 0 & P' & 0 & 0 & 0 & 0 & 0 & P_{\mathrm{d}} & 0 \\
0 & -P_{\mathrm{X}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 & 0 & 0 \\
0 & 0 & -P_{\mathrm{X}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 & 0 \\
P_{\mathrm{S}}^* & 0 & 0 & 0 & P'^* & 0 & 0 & 0 & -P_{\mathrm{U}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{U}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -P_3^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & P_{3\mathrm{d}} & 0 \\
0 & 0 & 0 & \sqrt{3}\,P_3^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\sqrt{3}\,P_{3\mathrm{d}} & 0 \\
0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & P_{\mathrm{d}}^* & 0 & 0 & 0 & 0 & P_{3\mathrm{d}}^* & -\sqrt{3}\,P_{3\mathrm{d}}^* & 0 & 0 & 0 & -P_{2\mathrm{d}} \\
0 & 0 & 0 & -P_2^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{2\mathrm{d}}^* & 0
\end{pmatrix} \tag{A.22}$$

$$\mathbf{P}_y = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & P_{\mathrm{S}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -P & 0 & 0 & 0 & 0 & -P_3 & -\sqrt{3}\,P_3 & 0 & 0 & 0 & -P_2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -P^* & 0 & 0 & 0 & 0 & P' & 0 & 0 & 0 & 0 & P_{\mathrm{d}} & 0 & 0 & 0 \\
0 & 0 & 0 & -P^*_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 \\
P^*_{\mathrm{S}} & 0 & 0 & 0 & P'^* & 0 & 0 & 0 & -P_{\mathrm{U}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -P^*_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -P^*_{\mathrm{U}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -P^*_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & P_{3\mathrm{d}} & 0 & 0 & 0 \\
0 & -\sqrt{3}\,P^*_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sqrt{3}\,P_{3\mathrm{d}} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & P^*_{\mathrm{d}} & 0 & 0 & 0 & 0 & P^*_{3\mathrm{d}} & \sqrt{3}\,P^*_{3\mathrm{d}} & 0 & 0 & 0 & -P_{2\mathrm{d}} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -P^*_{\mathrm{Xd}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -P^*_{\mathrm{Xd}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -P^*_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P^*_{2\mathrm{d}} & 0 & 0 & 0
\end{pmatrix} \tag{A.23}$$

$$\mathbf{P}_z = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & P_{\mathrm{S}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -P & 0 & 0 & 0 & 0 & 2\,P_3 & 0 & 0 & 0 & 0 & -P_2 \\
0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{X}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -P^* & 0 & 0 & P' & 0 & 0 & 0 & 0 & 0 & 0 & P_{\mathrm{d}} & 0 & 0 \\
P_{\mathrm{S}}^* & 0 & 0 & 0 & P'^* & 0 & 0 & 0 & -P_{\mathrm{U}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -P_{\mathrm{X}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 \\
0 & -P_{\mathrm{X}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -P_{\mathrm{U}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 2\,P_3^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2\,P_{3\mathrm{d}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & P_{\mathrm{d}}^* & 0 & 0 & 0 & 0 & -2\,P_{3\mathrm{d}}^* & 0 & 0 & 0 & 0 & -P_{2\mathrm{d}} \\
0 & 0 & 0 & 0 & 0 & 0 & -P_{\mathrm{Xd}}^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -P_2^* & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -P_{2\mathrm{d}}^* & 0 & 0
\end{pmatrix} \tag{A.24}$$

For the thirty-band model, we construct the momentum matrices as:

$$\mathbf{P}_{30} = \begin{bmatrix} \mathbf{P}_{15} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{15} \end{bmatrix} \tag{A.25}$$

where the subscript denotes the number of bands.

A.5 Spin-orbit matrix

This section describes the spin-orbit matrix $\mathbf{H}_{\mathrm{SO}}$ used in the thirty-band model introduced in Chapter 6. The spin splitting energies are defined in Section 6.1. We start with an all-zero 30×30 matrix, which is then modified as follows:

$$\mathbf{H}_{\mathrm{SO,2:4,17:19}} = \begin{pmatrix} 0 & 0 & \Delta\mathrm{i} \\ 0 & 0 & \Delta \\ \Delta^*\,\mathrm{i} & \Delta^* & 0 \end{pmatrix} \qquad \mathbf{H}_{\mathrm{SO,17:19,2:4}} = \begin{pmatrix} 0 & 0 & -\Delta\mathrm{i} \\ 0 & 0 & \Delta \\ -\Delta^*\,\mathrm{i} & \Delta^* & 0 \end{pmatrix}$$

$$\mathbf{H}_{\mathrm{SO,2:4,2:4}} = \begin{pmatrix} 0 & \Delta & 0 \\ \Delta^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{H}_{\mathrm{SO,17:19,17:19}} = \begin{pmatrix} 0 & -\Delta & 0 \\ -\Delta^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{H}_{\mathrm{SO,6:8,21:23}} = \begin{pmatrix} 0 & 0 & \Delta_{\mathrm{c}}\mathrm{i} \\ 0 & 0 & \Delta_{\mathrm{c}} \\ \Delta_{\mathrm{c}}^*\,\mathrm{i} & \Delta_{\mathrm{c}}^* & 0 \end{pmatrix} \qquad \mathbf{H}_{\mathrm{SO,21:23,6:8}} = \begin{pmatrix} 0 & 0 & -\Delta_{\mathrm{c}}\mathrm{i} \\ 0 & 0 & \Delta_{\mathrm{c}} \\ -\Delta_{\mathrm{c}}^*\,\mathrm{i} & \Delta_{\mathrm{c}}^* & 0 \end{pmatrix}$$

$$\mathbf{H}_{\mathrm{SO,6:8,6:8}} = \begin{pmatrix} 0 & \Delta_{\mathrm{c}} & 0 \\ \Delta_{\mathrm{c}}^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{H}_{\mathrm{SO,21:23,21:23}} = \begin{pmatrix} 0 & -\Delta_{\mathrm{c}} & 0 \\ -\Delta_{\mathrm{c}}^* & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{A.26}$$

where the subscripts denote ranges of row and column indices respectively.

A.6 Poisson's equation

For the solution of Poisson's equation in Chapter 7, we apply a FD discretization in both the x- and z-direction. Because we include the dielectric in the solution domain, N_z is now defined as the sum of the number of mesh points in the device and in the oxide: $N_z = N_{z,\mathrm{dev}} + N_{z,\mathrm{ox}}$. We then define the following

second order derivative matrices:

$$\mathbf{D2}_{x,\mathrm{p}} = \mathbf{D2}_x \otimes \mathbf{I}_z \tag{A.27}$$

$$\mathbf{D2}_{z,\mathrm{p}} = \mathbf{I}_x \otimes \mathbf{D2}_z \tag{A.28}$$

For the Gummel scheme, we define two extra terms: one for the left-hand side and one for the right-hand side of Poisson's equation:

$$\mathbf{G}_{\mathrm{LHS}} = -\frac{q}{\epsilon V_{\mathrm{ref}}}(\mathbf{n} + \mathbf{p}) \tag{A.29}$$

$$\mathbf{G}_{\mathrm{RHS}} = -\frac{q}{\epsilon V_{\mathrm{ref}}}(\mathbf{n} + \mathbf{p}) \circ \boldsymbol{\phi}_{\mathbf{old}} \tag{A.30}$$

where $\mathbf{n}$, $\mathbf{p}$ and $\boldsymbol{\phi}_{\mathbf{old}}$ are $N_x \cdot N_z$ vectors, representing the discretized versions of respectively the carrier densities of electrons and holes and the potential of the previous iteration. $\circ$ is the element-wise Hadamard product.

We then construct the left-hand side term of Poisson's equation in Eq. (7.26) as:

$$\mathbf{L} = \mathbf{D2}_{x,\mathrm{p}} + \mathbf{D2}_{z,\mathrm{p}} + \mathbf{G}_{\mathrm{LHS}} + \mathbf{B}_{\mathrm{N}} \tag{A.31}$$

where $\mathbf{B}_{\mathrm{N}}$ is an $N_x \cdot N_z \times N_x \cdot N_z$ matrix, in which the diagonal elements contain the Neumann boundary conditions applied to the corresponding mesh points. For the right-hand side, we have:

$$\mathbf{R} = -\frac{1}{\epsilon}\boldsymbol{\rho} + \mathbf{G}_{\mathrm{RHS}} - \mathbf{L}\mathbf{B}_{\mathrm{D}} \tag{A.32}$$

where $\mathbf{B}_{\mathrm{D}}$ is an $N_x \cdot N_z$ vector, containing the Dirichlet boundary values for the corresponding mesh points.

The new potential is then obtained from the solution of:

$$\boldsymbol{\phi}_{\mathbf{new}} = \mathbf{L} \backslash \mathbf{R} \tag{A.33}$$

Appendix B

Derivation spectral EF system

In this appendix, we derive the spectral EF system, based on the spectral decomposition discussed in Section 5.3.2. We first focus on the potential energy term, as its derivation is more involved than the rest of the EF system (Section B.1). We then insert the derived expression for the potential energy term into the full EF system (Section B.2). The derivation is carried out in full for the odd EFs. The differences for the even EFs are discussed at the end (Section B.3).

B.1 Potential energy term

We start with the potential energy term for the odd EFs. Inserting the sine expansion of the odd EFs of Eq. (5.3) into the potential energy term of the EF system of Eq. (3.15), we obtain:

$$\sum_{\mu'=-\infty}^{+\infty} \sum_{\mu=-\infty}^{+\infty} \tilde{V}_{\mathrm{e}}(x, k_{z\mu'}) \cos(k_{z\mu'} z) \tilde{F}_n^k(x, k_{z\mu}) \sin(k_{z\mu} z) \tag{B.1}$$

To be able to apply the completeness relation in the full system to obtain an equation for each spectral component, each term should only have a sine factor. We therefore rewrite the sine-cosine product in Eq. (B.1) as a difference of sines,

using a basic trigonometric identity, yielding for the potential energy term:

$$\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'})\tilde{F}_n^k(x,k_{z\mu})\left[\frac{\sin((k_{z\mu'}+k_{z\mu})z)-\sin((k_{z\mu'}-k_{z\mu})z)}{2}\right]$$

$$=\frac{1}{2}\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'})\tilde{F}_n^k(x,k_{z\mu})\sin((k_{z\mu'}+k_{z\mu})z)$$

$$-\frac{1}{2}\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'})\tilde{F}_n^k(x,k_{z\mu})\sin((k_{z\mu'}-k_{z\mu})z) \quad \text{(B.2)}$$

Next, we shift the variables from $k_{z\mu'}$ to $k_{z\mu'}-k_{z\mu}$ and $k_{z\mu'}+k_{z\mu}$ in respectively the first and the second term. This is possible because the summation index μ ranges from $-\infty$ to ∞ in Eqs. (5.3)-(5.4). This yields:

$$\frac{1}{2}\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z)$$

$$-\frac{1}{2}\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z)$$

$$=\frac{1}{2}\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z) \quad \text{(B.3)}$$

The sum over μ' can then be split up into a positive and a negative part:

$$\frac{1}{2}\sum_{\mu'=0}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z)$$

$$+\frac{1}{2}\sum_{\mu'=0}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,-k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,-k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(-k_{z\mu'}z) \quad \text{(B.4)}$$

$\tilde{V}_{\mathrm{e}}$ is the result of a cosine transform, and is therefore an even function, while conversely $\tilde{F}_n^k$ is the result of a sine transform, and is therefore odd. With these properties and the fact that the sine function itself is also odd, we rewrite the

second term of Eq. (B.4) as:

$$\sum_{\mu'=0}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z)$$

$$=2\sum_{\mu'=0}^{+\infty}\sum_{\mu=0}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z) \tag{B.5}$$

where we have split and recombined the sum over μ analogously to the sum over μ' in Eq. (B.4).

B.2 Full EF system

At this point, we turn to the full EF system and insert the spectral decomposition of the odd EFs (Eq. (5.3)) into the EF system of Eq. (3.15). The expression derived in Eq. (B.5) is inserted for the potential energy term. Working out the

derivatives in the rest of the system, we obtain:

$$
\begin{aligned}
&-\frac{\hbar^2}{2m_e}\sum_{\mu=0}^{+\infty}\frac{\partial^2\tilde{F}_n^k(x,k_{z\mu})}{\partial x^2}\sin(k_{z\mu}z)+k_y^2\frac{\hbar^2}{2m_e}\sum_{\mu=0}^{+\infty}\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu}z)\\
&+k_{z\mu}^2\frac{\hbar^2}{2m_e}\sum_{\mu=0}^{+\infty}\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu}z)-\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,x}(x)\sum_{\mu=0}^{+\infty}\frac{\partial\tilde{F}_m^k(x,k_{z\mu})}{\partial x}\sin(k_{z\mu}z)\\
&+k_y\frac{\hbar}{m_e}\sum_m p_{nm}^{k,y}(x)\sum_{\mu=0}^{+\infty}\tilde{F}_m^k(x,k_{z\mu})\sin(k_{z\mu}z)\\
&+\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x)\sum_{\mu=0}^{+\infty}k_{z\mu}\tilde{F}_m^k(x,k_{z\mu})\sin(k_{z\mu}z)\\
&+\sum_m H_{nm}^k(x)\sum_{\mu=0}^{+\infty}\tilde{F}_m^k(x,k_{z\mu})\sin(k_{z\mu}z)\\
&+\left[\sum_{\mu'=0}^{+\infty}\sum_{\mu=0}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu'}z)\right]\\
&=E\sum_{\mu=0}^{+\infty}\tilde{F}_n^k(x,k_{z\mu})\sin(k_{z\mu}z),
\end{aligned}
\tag{B.6}
$$

where the summation intervals of μ have been reduced to a one-sided summation starting from zero. Each term therefore acquires a factor of 2, which cancels at both sides of the equation. Note that Eq. (B.6) contains only terms with a sine function. To derive the final spectral EF system, we can therefore make use of the completeness relation of the sine set in each term, ending up with an

equation for each μ:

$$
\begin{aligned}
&-\frac{\hbar^2}{2m_e}\frac{\partial^2\tilde{F}_n^k(x,k_{z\mu})}{\partial x^2}+k_y^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu})+k_{z\mu}^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu})\\
&-\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,x}(x)\frac{\partial\tilde{F}_m^k(x,k_{z\mu})}{\partial x}+k_y\frac{\hbar}{m_e}\sum_m p_{nm}^{k,y}(x)\tilde{F}_m^k(x,k_{z\mu})\\
&+k_{z\mu}\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x)\tilde{F}_m^k(x,k_{z\mu})+\sum_m H_{nm}^k(x)\tilde{F}_m^k(x,k_{z\mu})\\
&+\sum_{\mu'=0}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})-\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu'})\\
&=E\tilde{F}_n^k(x,k_{z\mu}).
\end{aligned}
\tag{B.7}
$$

B.3 Even EFs

For the even EFs, the external potential term now contains a product of cosines:

$$
\sum_{\mu'=-\infty}^{+\infty}\sum_{\mu=-\infty}^{+\infty}\tilde{V}_{\mathrm{e}}(x,k_{z\mu'})\cos(k_{z\mu'}z)\tilde{F}_n^k(x,k_{z\mu})\cos(k_{z\mu}z) \tag{B.8}
$$

The derivation proceeds analogously to the odd case, now using the trigonometric relation:

$$
\cos(k_{z\mu'}z)\cos(k_{z\mu}z)=\left[\frac{\cos((k_{z\mu'}+k_{z\mu})z)+\cos((k_{z\mu'}-k_{z\mu})z)}{2}\right] \tag{B.9}
$$

resulting in the following spectral system for the even EFs:

$$
\begin{aligned}
&-\frac{\hbar^2}{2m_e}\frac{\partial^2\tilde{F}_n^k(x,k_{z\mu})}{\partial x^2}+k_y^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu})+k_{z\mu}^2\frac{\hbar^2}{2m_e}\tilde{F}_n^k(x,k_{z\mu})\\
&-\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,x}(x)\frac{\partial\tilde{F}_m^k(x,k_{z\mu})}{\partial x}+k_y\frac{\hbar}{m_e}\sum_m p_{nm}^{k,y}(x)\tilde{F}_m^k(x,k_{z\mu})\\
&-k_{z\mu}\frac{i\hbar}{m_e}\sum_m p_{nm}^{k,z}(x)\tilde{F}_m^k(x,k_{z\mu})+\sum_m H_{nm}^k(x)\tilde{F}_m^k(x,k_{z\mu})\\
&+\sum_{\mu'=0}^{+\infty}\left[\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}-k_{z\mu})+\tilde{V}_{\mathrm{e}}(x,k_{z\mu'}+k_{z\mu})\right]\tilde{F}_n^k(x,k_{z\mu'})\\
&=E\tilde{F}_n^k(x,k_{z\mu}).
\end{aligned}
\tag{B.10}
$$

This system differs with that of the odd EFs in two places. First, the potential term contains an autocorrelation instead of a convolution. Second, the term linear in $k_{z\mu}$ has a minus sign.

Appendix C

Pharos input files

C.1 Example input file

```
type = 'ivcurve2D';
global calc_real;
global calc_conc;
global save_partial;
global transformation;
constants;

%% Mesh definition
nox = 2000;
noz = 200;
nokz = 6;

polarity = 'nTFET'

halfwidth = 9.99*nm; % Half of Tbody
xloc_source = 20*nm; % Location of source contact
xloc_drain = -80*nm; % Location of drain contact

% Only used for self-consistency:
xloc_sourcechannel = 0*nm; % Location of source-channel interface
xloc_channeldrain = 50*nm; % Location of channel-drain interface
xloc_gate_1 = 0*nm; % Location of left end of gate
xloc_gate_2 = 20*nm; % Location of right end of gate

lsource = xloc_sourcechannel - xloc_source;
lchannel = xloc_channeldrain - xloc_sourcechannel;
ldrain = xloc_drain - xloc_channeldrain;
loversource = xloc_sourcechannel - xloc_gate_1;
loverdrain = xloc_gate_2 - xloc_channeldrain;
```

```
30
31  tox = 0.6*nm; % Physical oxide thickness
32  epsilon_ox = 3.9*perm_0; % Oxide dielectric constant
33
34  spatialmesh = devicegrid(linspace(xloc_drain,xloc_source,nox), ...
35      linspace(-halfwidth,halfwidth,noz), nokz)
36  doping = [];
37
38  %% Materials definition
39  load('ingaas30')
40  load('gasb30')
41  gasb_obj = gasb.copy;
42  ingaas_obj = ingaas.copy;
43
44  materials = [gasb_obj ingaas_obj];
45  basis_mat = 1; % Index in materials array, fastest if most prevalent material
46  xdiff = spatialmesh.xdiff;
47  zdiff = spatialmesh.zdiff;
48  materialgrid = [2*ones(noz,1+round(abs(xloc_drain)/xdiff)) ...
49      ones(noz,round(abs(xloc_source)/xdiff))];
50  epsilon_semicond = [materials(materialgrid).perm].';
51
52  %% Simulation settings
53  devicetype = 'TFET'; % Other option: 'diode', is used to speed up calculation for diodes
54  materialconf = 'simple_hetero'; % Heterostructure only in transport direction.
55                                  % Other options: 'simple_homo', 'complex'
56  strainconf = 'non-uniform' % Strain profile is non-uniform
57                             % Leave empty for unstrained or uniform strain
58  confinement = 1; % Device is confined in z-direction
59  transformation = 1 % Heterostructure transformation
60  calc_real = 0; % Calculate envelope functions in position domain (fastest if 0)
61  calc_conc = 0; % Calculate carrier density (fastest if 0, auto-activated for self-consistency)
62  save_partial = 1; % Save results after every bias point
63
64  %% Contact definition
65  contacts(1).type = 'source';
66  contacts(1).xloc = xloc_source;
67  contacts(1).zarray = linspace(-halfwidth,halfwidth,noz);
68  contacts(1).kzarray = pi/spatialmesh.Lz*[0:nokz-1];
69
70  contacts(2).type = 'drain';
71  contacts(2).xloc = xloc_drain;
72  contacts(2).zarray = linspace(-halfwidth,halfwidth,noz);
73  contacts(2).kzarray = pi/spatialmesh.Lz*[0:nokz-1];
74
75  injectioncontact = 0; % Specifies contact of injection as index in contacts array.
76                        % If 0, injection occurs in conduction band
77
78  %% Device instantiation
79  device = devicestruct(spatialmesh, lsource, ldrain, loversource, loverdrain, tox, doping, ...
80          materials, basis_mat, materialgrid, contacts, injectioncontact, devicetype, ...
81          materialconf, confinement);
82  device.strainconf = strainconf;
```

```
83  device.xlocsourcechannel = xloc_sourcechannel;
84  device.xlocchanneldrain = xloc_channeldrain;
85  device.lchannel = lchannel;
86
87  %% SDevice node definition and loading
88  switch halfwidth
89  %% 20nm
90     case 9.99*nm
91        disp('Tbody is 20nm')
92        SDevNodeNb = [654 655 656 657 658 659 660];  % Numbers of the SDevice nodes
93        Vgs = [0.7 0.8 0.9 1]*Volt;  % Gate-source voltage, only used in self-consistent run,
94                                     % otherwise just stored
95  %% 10nm
96     case 4.99*nm
97           disp('Tbody is 10nm')
98           SDevNodeNb = [120 121 128 122 129]; % Numbers of the SDevice nodes
99           Vgs = [0.7 0.8 0.9 1]*Volt; % Gate-source voltage, only used in self-consistent run,
100                                       % otherwise just stored
101 end
102 SDevNodeNb
103 fprintf('Number of bands: %i\n',materials(1).nobands)
104
105 sDevProject = '/imec/users/verreck/private/SentaurusDB/Non-uniform_stress/ ...
106       Non-uniform_strain_staggered_sourceunstrained/'; % Path of SDevice project
107 sDevNodes   = cellfun(@(c)  ['n' int2str(c)] ,num2cell(SDevNodeNb), ...
108       'UniformOutput', 0); % Convert node numbers into a cell
109 [V, ~, ~, Nd, Na, stressXX, stressYY, stressZZ, stressXY, stressXZ, stressYZ] = ...
110       sDeviceLoad2D(sDevProject, sDevNodes, 1, 1, 1); % Load potential, doping and stress
111 [strainXX, strainYY, strainZZ, strainXY, strainXZ, strainYZ] = ...
112       convertStresstoStrain2D(stressXX{1}, stressYY{1}, stressZZ{1}, ...
113       stressXY{1}, stressXZ{1}, stressYZ{1}, device);
114
115 %% Strain incorporation
116 meshlist = [kron(spatialmesh.xmesh.',ones(spatialmesh.noz,1)), ...
117       kron(ones(spatialmesh.nox,1),spatialmesh.zmesh.')];
118 e_tensor_list = [meshlist strainXX(meshlist(:,1), meshlist(:,2))
119       strainYY(meshlist(:,1), meshlist(:,2)) strainZZ(meshlist(:,1), meshlist(:,2)) ...
120       strainXY(meshlist(:,1), meshlist(:,2)) strainXZ(meshlist(:,1), meshlist(:,2)) ...
121       strainYZ(meshlist(:,1), meshlist(:,2))];
122 [materials, materialgrid, device] = incorporateStrain(e_tensor_list, device);
123
124 %% Drain-source bias
125 Vbias = 0.3*Volt*ones(1,length(SDevNodeNb)); % Drain-source bias, VDS
126 if(strcmp(polarity,'pTFET')) Vbias = -1*Vbias; end % Invert sign of bias if pTFET
127 noV = length(Vbias);
128 Va = [zeros(noV,1), Vbias.'];
129
130 %% Uncomment for self-consistency:
131 % PARALLEL = 1;
132 % [V] = selfconsistentPotential2D(device,epsilon_ox,epsilon_semicond, ...
133 %     V,Vbias,Vgs,Na,Nd,PARALLEL)
```

C.2 Parameter files

C.2.1 $In_{0.53}Ga_{0.47}As$

Two-band

```
ingaas = material('In_{.53}Ga_{.47}As');
ingaas.a = 5.8687e-10*m;
ingaas.perm = 13.9*perm_0;
ingaas.EA  = 4.5*eV;
ingaas.E   = [-0.74 0]*eV;
ingaas.Eg = 0.74*eV;
ingaas.Dso = 0.33*eV;
ingaas.m_eff = [ -0.45 0.041 ]*m_e; % Heavy hole orthogonal eff mass
%ingaas.m_eff = [ -0.052 0.043 ]*m_e; % Light hole orthogonal eff mass

P = EtoP(15*eV); % Interband momentum matrix element, Ep

ingaas.P{1} = [0 P ; -P 0];
ingaas.P{2} = [0 0 ;  0 0];
ingaas.P{3} = [0 0 ;  0 0];
ingaas.S = generateS(ingaas.P{1}, ingaas.a);
save('./input/materials/ingaas2', 'ingaas');
```

Fifteen-band

```
ingaas = material('In_{.53}Ga_{.47}As');
ingaas.a = 5.8687e-10*m;
ingaas.perm = 13.9*perm_0;
ingaas.EA  = 4.5*eV;
ingaas.Eg = 0.74*eV;
ingaas.E = refEc([-12.55 0.00 0.00 0.00 0.74 4.33 4.33 4.33 8.55*0.53+8.56*0.47 ...
                  9.88*0.53+10.17*0.47 9.88*0.53+10.17*0.47 11.89 11.89 11.89  ...
                  12.64*0.53+13.64*0.47])*eV;
ingaas.m_eff = [ -0.45 0.041 ]*m_e;

% Interband momentum matrix elements CMME  0.1784
P_ingaas.P    = EtoP(15*eV) ;
P_ingaas.P_X  = EtoP(16.09*eV);
P_ingaas.P_3  = EtoP( 4.49*eV);
P_ingaas.P_2  = EtoP( 7.5 *eV);
P_ingaas.P_S  = EtoP( 4.38*eV);
P_ingaas.P_d  = EtoP( 0.38 *eV);
P_ingaas.P_Xd = EtoP( 4.93*eV);
P_ingaas.P_3d = EtoP(10.8*eV);
P_ingaas.P_2d = EtoP( 16.44*eV);
P_ingaas.P_U  = EtoP(24.1*eV);
P_ingaas.Pp   = EtoP(0.12*eV);

```

```
ingaas.P = generateP15(P_ingaas);
save('./input/materials/ingaas15', 'ingaas');
```

Thirty-band

```
ingaas = material('In_{.53}Ga_{.47}As');
ingaas.a = 5.8687e-10*m;
ingaas.perm = 13.9*perm_0;
ingaas.EA  = 4.5*eV;
Dso = 0.33*eV; % Vurgaftman: interpol with bowing param, Perea (1980): direct meas.
Dso_c = 0*eV % For better hh-mass and E_X
Dso_p = 0.057*eV; % Saidi, linear interpolation
E = refEc([-12.55*eV -Dso/3 -Dso/3 -Dso/3 0.74*eV 4.33*eV+2*Dso_c/3 ...
      4.33*eV+2*Dso_c/3 4.33*eV+2*Dso_c/3 (8.55*0.53+8.56*0.47)*eV ...
      (9.88*0.53+10.17*0.47)*eV (9.88*0.53+10.17*0.47)*eV 11.89*eV 11.89*eV 11.89*eV
       (12.64*0.53+13.64*0.47)*eV]);
ingaas.E = kron([1 1], E);
ingaas.Eg = 0.74*eV;
ingaas.m_eff = [ -0.45 0.041 ]*m_e;

% Interband momentum matrix elements, CMME = 0.0194
P_ingaasc.P    = -0.596000000000000004000*1i;
P_ingaasc.P_X  =  -0.50644377218047598000*1i;
P_ingaasc.P_3  =  -0.24305534958963951000*1i;
P_ingaasc.P_2  =  -0.00003712302712594492*1i;
P_ingaasc.P_S  =  -0.03037861775670625300*1i;
P_ingaasc.P_d  =  -0.03379750947028648700*1i;
P_ingaasc.P_Xd =  -0.323893135616685090000*1i;
P_ingaasc.P_3d =  -0.38004448270399599000*1i;
P_ingaasc.P_2d =  -0.62660381554365430000*1i;
P_ingaasc.P_U  =  -0.60039199152554534000*1i;
P_ingaasc.Pp   = EtoP(0.01*eV);

P15 = generateP15(P_ingaasc);
S15 = generateS(P15, ingaas.a);
P30{1} = kron(eye(2,2), P15{1});
P30{2} = kron(eye(2,2), P15{2});
P30{3} = kron(eye(2,2), P15{3});
ingaas.P = P30;
ingaas.S = kron(eye(2,2), S15);
ingaas.SO = generateSO30(-1i*Dso/3, -1i*Dso_c/3, -1i*Dso_p/3);

% Deformation potential constants
ingaas.a_c = 6.71*eV; % Vurgaftman, interpolation with bowing param
ingaas.a_v = -1.08*eV; % Vurgaftman, linear interpolation
ingaas.b_v = 1.89*eV; % Vurgaftman, linear interpolation
ingaas.d_v = -4.16*eV; % Vurgaftman, linear interpolation

% Elastic stiffness constant in Pa (Ioffe)
ingaas.C11 = 1*10^11;
ingaas.C12 = 4.92*10^10;
ingaas.C44 = 4.89*10^10;
```

```

% Elements of strain tensor for uniform normal strain
e_par = 0;
e_perp = 0;
ingaas.e_tensor = [e_perp e_par e_par 0 0 0];

save('./input/materials/ingaas30', 'ingaas');
```

C.2.2 $GaAs_{0.5}Sb_{0.5}$

Two-band

```
gaassb = material('Ga_As_{.5}Sb_{.5}');
gaassb.a = 5.8687e-10*m;
gaassb.perm = 14.3*perm_0;
gaassb.EA  = 4.07*eV;
gaassb.E   = [-0.71 0]*eV;
gaassb.Eg = 0.71*eV;
gaassb.Dso = 0.4*eV; % Quadratic interpolation between GaAs and GaSb
gaassb.m_eff = [ -0.455 0.045 ]*m_e;

P = EtoP(15.07*eV); % Interband momentum matrix element, Ep

gaassb.P{1} = [0 P ; -P 0];
gaassb.P{2} = [0 0 ;  0 0];
gaassb.P{3} = [0 0 ;  0 0];
gaassb.S = generateS(gaassb.P{1}, gaassb.a);
save('./input/materials/gaas_05_sb2', 'gaassb');
```

Fifteen-band

```
gaassb = material('Ga_As_{.5}Sb_{.5}');
gaassb.a = 5.8687e-10*m;
gaassb.perm = 14.3*perm_0;
gaassb.EA  = 4.07*eV;
gaassb.E = refEc([-12.17 0 0 0 0.72 3.89 3.89 3.89 8.56 9.59 10.17 ...
      10.99 10.99 10.99 13.19])*eV;
gaassb.m_eff = [ -0.455 0.045 ]*m_e;

% Interband momentum matrix elements
P_gaassb.P    = EtoP(15.8*eV);
P_gaassb.P_X  = EtoP(15.47*eV);
P_gaassb.P_3  = EtoP( 5.61 *eV  );
P_gaassb.P_2  = EtoP( 0.125 *eV );
P_gaassb.P_S  = EtoP( 5.97 *eV );
P_gaassb.P_d  = EtoP( 1.2 *eV);
P_gaassb.P_Xd = EtoP( 5.15 *eV );
P_gaassb.P_3d = EtoP( 12.19 *eV );
P_gaassb.P_2d = EtoP(10.95 *eV );
```

```
P_gaassb.P_U  = EtoP(25.55 *eV);
P_gaassb.Pp   = EtoP( 0.09*eV);

gaassb.P = generateP15(P_gaassb);
gaassb.S = generateS(gaassb.P, gaassb.a);
save('./input/materials/gaas_05_sb15', 'gaassb');
```

Thirty-band

```
gaassb= material('Ga_As_{.5}Sb_{.5}');
gaassb.a = 5.8687e-10*m;
gaassb.perm = 14.3*perm_0;
gaassb.EA  = 4.07*eV;
Dso = 0.4*eV;   % % Split-off energy, Vurgaftman: interpolation with bowing param
Dso_c =  0.18*eV; % GaAs Saidi;
Dso_p =  0.17*eV; % GaAs Saidi;
E = refEc([-12.17*eV -Dso/3 -Dso/3 -Dso/3  0.72*eV ...
      3.89*eV+2*Dso_c/3 3.89*eV+2*Dso_c/3 3.89*eV+2*Dso_c/3 8.56*eV 9.59*eV ...
      10.17*eV 10.99*eV 10.99*eV 10.99*eV 13.19*eV]);
gaassb.E = kron([1 1], E);
gaassb.Eg = 0.72*eV;
gaassb.m_eff = [ -0.455 0.045 ]*m_e;

% Interband momentum matrix elements, CMME 0.0352
P_gaassbc.P    = -0.560301952921453*1i;
P_gaassbc.P_X  =  -0.488843662185158*1i;
P_gaassbc.P_3  =  -0.234705103862240*1i;
P_gaassbc.P_2  =                  0*1i;
P_gaassbc.P_S  =  -0.047272452537300*1i;
P_gaassbc.P_d  =  -0.144568270258876*1i;
P_gaassbc.P_Xd =  -0.262713675777191*1i;
P_gaassbc.P_3d =  -0.437396800511563*1i;
P_gaassbc.P_2d =  -0.668523660965734*1i;
P_gaassbc.P_U  =  -0.545545870651279*1i;
P_gaassbc.Pp   =  -0.001455888380776*1i;

P15 = generateP15(P_gaassbc);
S15 = generateS(P15, gaassb.a);
P30{1} = kron(eye(2,2), P15{1});
P30{2} = kron(eye(2,2), P15{2});
P30{3} = kron(eye(2,2), P15{3});
gaassb.P = P30;
gaassb.S = kron(eye(2,2), S15);
gaassb.SO = generateSO30(-1i*Dso/3, -1i*Dso_c/3, -1i*Dso_p/3);

% Deformation potential constants
gaassb.a_c = 7.33*eV; % Vurgaftman, linear interpolation
gaassb.a_v = -0.98*eV; % Vurgaftman, linear interpolation
gaassb.b_v = 2*eV; % Vurgaftman, linear interpolation
gaassb.d_v = -4.75*eV; % Vurgaftman, linear interpolation

% Elastic stiffness constant in Pa (Ioffe)
```

```
gaassb.C11 = 1*10^11; % elastic stiffness constant in Pa (Ioffe)
gaassb.C12 = 4.68*10^10;
gaassb.C44 = 5.14*10^10;

% Elements of strain tensor for uniform normal strain
e_par = 0;
e_perp = 0;
gaassb.e_tensor = [e_perp e_par e_par 0 0 0];

save('./input/materials/gaas_05_sb30', 'gaassb');
```

C.2.3 InP

Two-band

```
inp = material('InP');
inp.a = 5.8687e-10*m;
inp.perm = 12.5*perm_0;
inp.EA  = 4.38*eV;
inp.E   = [-1.344 0]*eV;
inp.Eg = 1.344*eV;
inp.Dso = 0.108*eV;
inp.m_eff = [ -0.6 0.08 ]*m_e;

p = EtoP(15.46*eV); % Interband momentum matrix element, Ep

inp.P{1} = [0 p ; -p 0];
inp.P{2} = [0 0 ;  0 0];
inp.P{3} = [0 0 ;  0 0];
inp.S = generateS(inp.P{1}, inp.a);
save('./input/materials/inp2', 'inp');
```

C.2.4 GaSb

Fifteen-band

```
gasb = material('GaSb');
gasb.a  = 6.096e-10*m;
gasb.perm = 15.7*perm_0;
gasb.EA  = 4.06*eV;
gasb.E = refEc([-12.13 0 0 0 0.73 3.82 3.82 3.82 8.56 9.53 10.17 10.89 10.89 10.89 13.14])*eV;
gasb.Eg = 0.73*eV;

% Interband momentum matrix elements
P_gasb.P    =  EtoP(18  *eV);
P_gasb.P_X  =  EtoP(15.15  *eV) ;
P_gasb.P_3  =  EtoP(5.00 *eV);
P_gasb.P_2  =  EtoP( 0.14 *eV );
```

```
P_gasb.P_S  =  EtoP( 5.8 *eV );
P_gasb.P_d  =  EtoP( 0.81 *eV );
P_gasb.P_Xd =  EtoP( 5.35 *eV );
P_gasb.P_3d =  EtoP(10.99 *eV );
P_gasb.P_2d =  EtoP(13.96  *eV);
P_gasb.P_U  =  EtoP(26.2  *eV);
P_gasb.Pp   =  EtoP( 0.07*eV);

gasb.P = generateP15(P_gasb);
gasb.S = generateS(gasb.P, gasb.a);
save('./input/materials/gasb15', 'gasb');
```

Thirty-band

```
gasb = material('GaSb');
gasb.a  = 6.096*ang;
gasb.perm = 15.7*perm_0;
gasb.EA  = 4.06*eV;
Dso = 0.8*eV;   % Ioffe
Dso_c =  0*eV;
Dso_p =  0*eV;
E = refEc([-12.13*eV -Dso/3 -Dso/3 -Dso/3 0.73*eV 3.82*eV+2*Dso_c/3 3.82*eV+2*Dso_c/3 ...
      3.82*eV+2*Dso_c/3 8.56*eV 9.53*eV 10.17*eV 10.89*eV 10.89*eV 10.89*eV 13.14*eV]);
gasb.E = kron([1 1], E);
gasb.Eg = 0.73*eV;
gasb.m_eff = [ -0.4 0.041 ]*m_e;

P_gasbc.P = -0.61810763963209947000*1i; % CMME 0.2387
P_gasbc.P_X  =  -0.49568600493041387000*1i;
P_gasbc.P_3  =  -0.30831240529344783000*1i ;
P_gasbc.P_2  =  -0.02424798202633201800*1i ;
P_gasbc.P_S  =  -0.31010652372414255000*1i ;
P_gasbc.P_d  =  -0.08292524499885280100*1i ;
P_gasbc.P_Xd =  -0.30520939426029781000*1i ;
P_gasbc.P_3d =  -0.38005746803848456000*1i ;
P_gasbc.P_2d =  -0.50656927714354105000*1i ;
P_gasbc.P_U  =  -0.69360404289505151000*1i ;
P_gasbc.Pp   =  -0.01769152476027870800*1i ;

P15 = generateP15(P_gasbc);
S15 = generateS(P15, gasb.a);
P30{1} = kron(eye(2,2), P15{1});
P30{2} = kron(eye(2,2), P15{2});
P30{3} = kron(eye(2,2), P15{3});
gasb.P = P30;
gasb.S = kron(eye(2,2), S15);

% Deformation potential constants
gasb.a_c = 7.5*eV;
gasb.a_v = -0.8*eV;
gasb.b_v = 2.0*eV;
gasb.d_v = -4.7*eV;
```

```
39
40 % Elastic stiffness constants in Pa (Ioffe)
41 gasb.C11 = 8.83*10^10; % elastic stiffness constant in Pa (Ioffe)
42 gasb.C12 = 4.02*10^10;
43 gasb.C44 = 4.32*10^10;
44
45 % Elements of strain tensor for uniform normal strain
46 e_par = 0;
47 e_perp = 0;
48 gasb.e_tensor = [e_par e_par e_perp 0 0 0];
49 gasb.SO = generateSO30(-1i*Dso/3, -1i*Dso_c/3, -1i*Dso_p/3);
50 save('./input/materials/gasb30', 'gasb');
```

C.2.5 InAs

Fifteen-band

```
1  inas = material('InAs');
2  inas.a  = 6.0583*ang;
3  inas.perm = 15.15*perm_0;
4  inas.EA  = 4.9*eV;
5  inas.E = refEc([-12.69 0.00 0.00 0.00 0.37 4.39 4.39 4.39 8.55 ...
6        9.88 9.88 11.89 11.89 11.89 12.64])*eV;
7  inas.Eg = 0.37*eV;
8  inas.m_eff = [ -0.41 0.023 ]*m_e;
9
10 % Interband momentum matrix elements
11 P_inas.P    = EtoP(19.04*eV);
12 P_inas.P_X  = EtoP(15.64*eV);
13 P_inas.P_3  = EtoP( 3.89*eV);
14 P_inas.P_2  = EtoP( 1.00*eV);
15 P_inas.P_S  = EtoP( 5.00*eV);
16 P_inas.P_d  = EtoP( 0.1 *eV);
17 P_inas.P_Xd = EtoP( 5.00*eV);
18 P_inas.P_3d = EtoP(11.66*eV);
19 P_inas.P_2d = EtoP( 2.50*eV);
20 P_inas.P_U  = EtoP(19.00*eV);
21 P_inas.Pp   = EtoP(0.01*eV);
22
23 inas.P = generateP15(P_inas);
24 inas.S = generateS(inas.P, inas.a);
25 save('./input/materials/inas15', 'inas');
```

Thirty-band

```
1  inas = material('InAs');
2  inas.a  = 6.0583*ang;
3  inas.perm = 15.15*perm_0;
4  inas.EA  = 4.9*eV;
```

```
Dso = 0.43*eV;
Dso_c = 0.24*eV;
Dso_p = 0.18*eV;
E = refEc([-12.69*eV -Dso/3 -Dso/3 -Dso/3 0.354*eV ...
      4.39*eV+2*Dso_c/3  4.39*eV+2*Dso_c/3 4.39*eV+2*Dso_c/3 8.55*eV 9.88*eV ...
      9.88*eV 11.89*eV 11.89*eV 11.89*eV 12.64*eV]);
inas.E = kron([1 1], E);
inas.Eg = 0.354*eV;
inas.m_eff = [ -0.41 0.023 ]*m_e;

% Interband momentum matrix elements
P_inas.P    = EtoP(19.04*eV);
P_inas.P_X  = EtoP(15.64*eV);
P_inas.P_3  = EtoP( 3.89*eV);
P_inas.P_2  = EtoP( 1.00*eV);
P_inas.P_S  = EtoP( 5.00*eV);
P_inas.P_d  = EtoP( 0.1 *eV);
P_inas.P_Xd = EtoP( 5.00*eV);
P_inas.P_3d = EtoP(11.66*eV);
P_inas.P_2d = EtoP( 2.50*eV);
P_inas.P_U  = EtoP(19.00*eV);
P_inas.Pp   = EtoP(0.01*eV);

P15 = generateP15(P_inas);
P30{1} = kron(eye(2,2), P15{1});
P30{2} = kron(eye(2,2), P15{2});
P30{3} = kron(eye(2,2), P15{3});
inas.P = P30;
S15 = generateS(inas.P, inas.a);
inas.S = generateS(inas.P, inas.a);
inas.SO = generateSO30(-1i*Dso/3, -1i*Dso_c/3, -1i*Dso_p/3);

% Deformation potential constants
inas.a_c = 5.08*eV; % Vurgaftman
inas.a_v = -1*eV; % Vurgaftman
inas.b_v = 1.8*eV; % Vurgaftman
inas.d_v = -3.6*eV; % Vurgaftman

% Elastic stiffness constants in Pa (Ioffe)
inas.C11 = 8.34e10;
inas.C12 = 4.54e10;
inas.C44 = 3.95e10;

% Elements of strain tensor for uniform normal strain
e_perp = 0;
e_par = 0;
inas.e_tensor = [e_par e_par e_perp 0 0 0];

save('./input/materials/inas30', 'inas');
```

Appendix D

Sentaurus Device input files

D.1 SDEditor file pocketed pointTFET

This section contains the SDevice Editor input file of a pocketed pointTFET. The source and drain regions are defined to be longer than in the QM simulator. When transferring the electrostatic potential to the QM simulator, these regions are cut off where the potential becomes flat in the transport direction.

An ellipsis ("...") indicates a line of code has been split over two page lines. Before running the code, these lines should be reconnected on one line with the ellipsis removed.

```
1  ;--- Define parameters
2  (define X 0.050)   ;# Channel length
3  (define Y @HalfBodyT@)   ;# Half body thickness
4  (define NSource @NSource@)   ;# Source doping concentration
5  (define NDrain @NDrain@)   ;# Drain doping concentration
6  (define GateL @GateL@)   ;# Gate length
7  (define SOverlap @SOverlap@)   ;# Overlap of gate over source
8  (define Tox 0.0006)   ;# Physical oxide thickness
9  (define Tpo @Tpo@)   ;# Pocket thickness
10 (define Npo @Npo@)   ;# Pocket doping concentration
11 (define Tpo_2 @Tpo_2@)   ;# Source pocket thickness
12 (define Npo_2 @Npo_2@)   ;# Source pocket doping concentration
13
14 ;--- Create regions
15 (isegeo:create-rectangle (position (* -1 X) (* -1 Y) 0) ...
16       (position 0 Y 0) "@Material1@" "region_1") ;# Source region
17 (isegeo:set-default-boolean "ABA")
18 (isegeo:create-rectangle (position X (* -1 Y) 0) ...
```

```
19        (position 0 Y 0) "@Material2@" "region_2" ) ;# Channel region
20    (isegeo:create-rectangle (position X (* -1 Y) 0) ...
21        (position (* 2 X) Y 0) "@Material2@" "region_3") ;# Drain region
22    (isegeo:create-rectangle (position (* -1 X) (+ (*  1 Y) Tox) 0) ...
23        (position (* 2 X) (*  1 Y) 0) "Insulator1" "region_4")
24    (isegeo:create-rectangle (position (* -1 X) (- (* -1 Y) Tox) 0) ...
25        (position (* 2 X) (* -1 Y) 0) "Insulator1" "region_4") ;# Oxide
26    (if (> Tpo 0) ;# Pocket
27    (begin
28    (isegeo:create-rectangle (position Tpo (* -1 Y) 0) ...
29        (position 0 (+ Y) 0) "@Material2@" "region_5")
30    ))
31    (if (> Tpo_2 0) ;# Source pocket
32    (begin
33    (isegeo:create-rectangle (position 0 (* -1 Y) 0) ...
34        (position (* -1 Tpo_2) Y 0) "@Material2@" "region_6")
35    ))
36    (sdegeo:insert-vertex (position (- GateL SOverlap) (+ Y Tox) 0))
37    (sdegeo:insert-vertex (position (- GateL SOverlap) (- (* -1 Y) Tox) 0))
38    (sdegeo:insert-vertex (position (- 0 SOverlap) (+ Y Tox) 0))
39    (sdegeo:insert-vertex (position (- 0 SOverlap) (- (* -1 Y) Tox) 0))
40
41    ;--- Doping
42    (isedr:define-constant-profile "region_1_doping" "BoronActiveConcentration" NSource)
43    (isedr:define-constant-profile-region "region_1_doping"        "region_1_doping" "region_1")
44    (isedr:define-constant-profile "region_2_doping" "ArsenicActiveConcentration" 1e10)
45    (isedr:define-constant-profile-region "region_2_doping"        "region_2_doping" "region_2")
46    (isedr:define-constant-profile "region_3_doping" "ArsenicActiveConcentration" NDrain)
47    (isedr:define-constant-profile-region "region_3_doping"        "region_3_doping" "region_3")
48    (if (> Tpo 0)
49    (begin
50    (isedr:define-constant-profile "region_5_doping" "ArsenicActiveConcentration" Npo)
51    (isedr:define-constant-profile-region "region_5_doping"        "region_5_doping" "region_5")
52    ))
53    (if (> Tpo_2 0)
54    (begin
55    (isedr:define-constant-profile "region_6_doping" "BoronActiveConcentration" Npo_2)
56    (isedr:define-constant-profile-region "region_6_doping" "region_6_doping" "region_6")
57    ))
58
59    ;--- Meshing
60    (sdedr:define-refinement-window "a" "Rectangle" (position (* -1 X)  (* -1 Y)  0) (position 0 Y 0))
61    (sdedr:define-multibox-size "b" 0.005 0.005 0.0001 0.0001 -1.1 -1.1)
62    (sdedr:define-multibox-placement "y" "b" "a")
63    (sdedr:define-refinement-window "w" "Rectangle" (position X  (* -1 Y)  0) (position 0 Y 0))
64    (sdedr:define-multibox-size "c" 0.005 0.005 0.0001 0.0001  1.1 -1.1)
65    (sdedr:define-multibox-placement "z" "c" "w")
66    (if (> Tpo 0)
67    (begin
68    (sdedr:define-refinement-window "p" "Rectangle" (position (* 1 Tpo) (* -1 Y) 0) ...
69        (position (* -1 Tpo) Y 0))
70    (sdedr:define-multibox-size "o" 0.005 0.005 0.0001 0.0001 -1 -1)
71    (sdedr:define-multibox-placement "q" "o" "p")
```

```
))
(sdedr:define-refinement-window "a1" "Rectangle" (position (* -1 X) (* -1 Y) 0) (position 0 Y 0))
(sdedr:define-multibox-size "b1" 0.005 0.005 0.0001 0.0001 -1.1 1.1)
(sdedr:define-multibox-placement "y1" "b1" "a1")
(sdedr:define-refinement-window "w1" "Rectangle" (position X (* -1 Y) 0) (position 0 Y 0))
(sdedr:define-multibox-size "c1" 0.005 0.005 0.0001 0.0001 1.1 1.1)
(sdedr:define-multibox-placement "z1" "c1" "w1")
(sdedr:define-refinement-window "w2" "Rectangle" (position X (* -1 Y) 0) (position 0 Y 0))
(sdedr:define-multibox-size "c2" 0.005 0.005 0.0001 0.0001 -1.1 1.1)
(sdedr:define-multibox-placement "z2" "c2" "w2")
(sdedr:define-refinement-window "Ref.Win" "Rectangle" (position -0.20 -0.4 0.0) ...
      (position 0.3 0.4 0.0))
(sdedr:define-refinement-size "Ref.Def1" 0.001 0.001 0.001 0.001)
(sdedr:define-refinement-placement "Doping.Pl" "Ref.Def1" "Ref.Win")

;--- Contacts
(sdegeo:define-contact-set "Gate"   4  (color:rgb 1 0 0 ) "##" )
(sdegeo:define-contact-set "Source" 4  (color:rgb 1 1 0 ) "##" )
(sdegeo:define-contact-set "Drain"  4  (color:rgb 1 0 1 ) "##" )
(sdegeo:set-current-contact-set "Gate")
(sdegeo:define-2d-contact (list (car (find-edge-id (position (+ 0.001 (- 0 SOverlap)) ...
      (+ Y Tox) 0)))) "Gate")
(sdegeo:define-2d-contact (list (car (find-edge-id (position (+ 0.001 (- 0 SOverlap)) ...
      (- (* -1 Y) Tox) 0)))) "Gate")
(sdegeo:set-current-contact-set "Source")
(sdegeo:define-2d-contact (list (car (find-edge-id (position (* -1 X) (* 0 Y) 0)))) "Source")
(sdegeo:set-current-contact-set "Drain")
(sdegeo:define-2d-contact (list (car (find-edge-id (position (* 2 X) (* 0 Y) 0)))) "Drain")

;--- Saving
(sdeio:save-tdr-bnd (get-body-list) "n@node@_bnd.tdr")
(sde:build-mesh "mesh" "-F tdr -s" "n@node@_msh")

```

D.2 Command file

This section contains an SDevice command file, which is used to calculate the electrostatic potential in the configurations defined in the previous section. The same command file is used for semi-classical transport calculations, by uncommenting the lines indicated below.

```
File{
  Grid= "@tdr@"
  parameter="@parameter@"
  Plot= "@tdrdat@"
  Current= "@plot@"
}

Electrode {
```

```
  { Name="Source"    Voltage=0  }
  { Name="Drain"     Voltage=@Vdrain@  }
  { Name="Gate"      Voltage=@Vgate@     Workfunction=5 }
}

Physics{
  Fermi
  eMultivalley(NonParabolicity)  hMultivalley(NonParabolicity)
  EffectiveIntrinsicDensity (NoBandGapNarrowing)
# Uncomment for transport simulations:
      # Mobility (DopingDependence)
      # Mobility (eHighFieldSaturation(Eparallel))
      # Recombination(
      #     Band2Band(Model=NonlocalPath1
      #     -InterfaceReflection
      #     -FranzDispersion ))
}

Plot{
  eDensity hDensity eCurrent hCurrent TotalCurrent TotalCurrentDensity
  eBand2BandGeneration hBand2BandGeneration Doping DonorConcentration AcceptorConcentration
  ConductionBand ValenceBand eQuasiFermi hQuasiFermi Potential ElectricField/Vector
  BandGap EffectiveBandGap SemiconductorGradValenceBand/Vector DielectricConstant ElectronAffinity
}

Math{
  Extrapolate RelErrControl Currentweighting
  Iterations=50
  NotDamped=50
  Digits=5
  RhsFactor = 1e25
}

Solve {
  Coupled {Poisson}
  Coupled {Poisson Electron Hole}
# Uncomment for transport simulations:
      #Quasistationary(
      #   InitialStep=1e-3 Increment=2
      #   MinStep=1e-7 MaxStep=0.1
      #   Goal { Name="Gate" Voltage=@Vgate@}
      #){Coupled {Poisson Electron Hole} }
}
```

D.3 Parameter files

This section contains the material parameter files for the SDevice simulations. The parameters have been obtained from Levinshtein *et al.* [81] and Vurgaftman *et al.* [139].

Material names which are not natively present in SDevice, can be added by including the following lines in datexcodes.txt, which is located in the project folder, e.g. for GaAsSb:

```
DATEX2.1
Datacode
Materials {
GaAsSb {
label = "GaAsSb"
group = Semiconductor
color = #8298d9, #93a9ea  }}
```

D.3.1 Oxide

```
Material = "Oxide" {
Bandgap
{ * Eg = Eg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
      Chi0 = 2.05     # [eV] electron affinity
      Eg0  = 5.9      # [eV]
      alpha = 0.0000e+00      # [eV K^-1]
      beta  = 0.0000e+00      # [K]
      Tpar  = 0.0000e+00      # [K]
}
Epsilon{epsilon = 3.9}}
```

D.3.2 $In_{0.53}Ga_{0.47}As$

```
Material = "In0.53Ga0.47As" {
Bandgap{
  * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
      Chi0       = 4.5       # [eV] electron affinity
      Bgn2Chi  = 0.5       # [1]
      Eg0  = 0.74     # [eV]
      alpha      = 0    # [eV K^-1]
      beta       = 6.3600e+02  # [K]
      Tpar       = 0.0000e+00  # [K]
}
Band2BandTunneling{
      #Apath1   = 1.250e+20    # [1/cm^3/sec] # used for lineTFET
      #Bpath1   = 4.3800e+06  # [V/cm]
      Apath1    = 2.045e+20    # [1/cm^3/sec] # calibrated values, hh
      Bpath1    = 5.7000e+06  # [V/cm]
      Dpath1    = 0.0000e+00  # [eV]
      Ppath1    = 0.0             # [eV]
      Rpath1    = 0.0000e+00  # [1]
      MaxTunnelLength    = 1e-05    # [cm]
```

```
}
MultiValley{
# Used for two-band and fifteen-band simulations
  # eValley"Gamma"(m0 = 0.041 energy = 0 degeneracy = 1 alpha0 = 1.24) #[1,eV,1,eV^-1]
  # eValley"Cond2"(m0 = 0.29 energy = 0.46 degeneracy = 4 alpha0 = 0.42)
  # eValley"Cond3"(m0 = 0.68 energy = 0.59 degeneracy = 3 alpha0 = 0.077)
  # hValley"LH"(m0 = 0.052 energy = 0 degeneracy = 1 alpha0 = 0.0)
  # hValley"HH"(m0 = 0.450 energy = 0 degeneracy = 1 alpha0 = 0.0)
  # hValley"SO"(m0 = 0.150 energy = -0.33 degeneracy = 1 alpha0 = 0.0)
# Used for thirty-band simulations
  eValley"Gamma"(m0 = 0.0409 energy = 0 degeneracy = 1 alpha0 = 1.24 xid = -6.71)  #[1,eV,1,eV^-1]
  eValley"Cond2"(m0 = 0.29 energy = 0.46 degeneracy = 4 alpha0 = 0.42)
  eValley"Cond3"(m0 = 0.68 energy = 0.59 degeneracy = 3 alpha0 = 0.077)
  hValley"LH"(m0 = 0.0545 energy = 0 degeneracy = 1 alpha0 = 0.0)
  hValley"HH"(m0 = 0.44970 energy = 0 degeneracy = 1 alpha0 = 0.0)
  hValley"SO"(m0 = 0.1350 energy = -0.33 degeneracy = 1 alpha0 = 0.0)
}
Epsilon{epsilon       = 13.9}    # [1]
}
```

D.3.3 $GaAs_{0.5}Sb_{0.5}$

```
Material="GaAsSb"{
Bandgap{
 * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
 * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
     Chi0       = 4.07      # [eV] electron affinity
     Bgn2Chi  = 0.5       # [1]
     Eg0  = 0.72      # [eV]
     alpha      = 0  # [eV K^-1]
     beta       = 6.3600e+02  # [K]
     Tpar       = 0.0000e+00  # [K]
}
Band2BandTunneling{
   Apath1       = 1.7300e+20  # [1/cm^3/sec] ; based on p-values: 1.26e+20
     Bpath1     = 5.580e+06   # [V/cm] 4.08e+06
     Dpath1     = 0.0000e+00  # [eV]
     Ppath1     = 0.0         # [eV]
     Rpath1     = 0.0000e+00  # [1]
     MaxTunnelLength    = 1e-05    # [cm]
}
MultiValley{
     # Used in two-band simulations:
     # eValley"Gamma"(m0 = 0.045 energy = 0 degeneracy = 1 alpha0 = 1.28)    #[1,eV,1,eV^-1]
     # hValley"LH"(m0 = 0.066 energy = 0 degeneracy = 1 alpha0 = 0.0)
     # hValley"HH"(m0 = 0.455 energy = 0 degeneracy =  1 alpha0 = 0.0)

     eValley(m0 = 0.0459 energy = 0 degeneracy = 1 alpha0 = 1.28)    #[1,eV,1,eV^-1]
     hValley(m0 = 0.0598 energy = 0 degeneracy = 1 alpha0 = 0.0)
     hValley(m0 = 0.455  energy = 0 degeneracy = 1 alpha0 = 0.0)
     hValley(m0 = 0.1547 energy = -0.4 degeneracy = 1 alpha0 = 0.0)
```

```
}
Epsilon{epsilon       = 14.3}     # [1]
}
```

D.3.4 InP

```
Material="InP"{
Bandgap{
  * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
       Chi0        = 4.38      # [eV] electron affinity
       Bgn2Chi  = 0.5       # [1]
       Eg0  = 1.344    # [eV]
       alpha       = 0 # [eV K^-1]
       beta        = 6.3600e+02  # [K]
       Tpar        = 0.0000e+00  # [K]
}
Band2BandTunneling{
     Apath1       = 1.3800e+20  # [1/cm^3/sec]
       Bpath1     = 1.60e+07      # [V/cm]
       Dpath1     = 0.0000e+00  # [eV]
       Ppath1     = 0.0               # [eV]
       Rpath1     = 0.0000e+00  # [1]
       MaxTunnelLength    = 1e-05    # [cm]
}
MultiValley{
       eValley(0.08, 0.08, 0.08, 0.0000e+00, 1, 0.63)        #[1,1,1,eV,1,eV^-1]
       eValley(0.25, 0.25, 0.25, 0.59, 4, 0.29)
       hValley(0.089, 0.089, 0.089, 0.0000e+00, 1, 0.0)
       hValley(0.60, 0.60, 0.60, 0.0000e+00, 1, 0.0)
}
Epsilon{epsilon       = 12.5}     # [1]
}
```

D.3.5 GaSb

```
Material="GaSb"{
Bandgap{
  * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
       Chi0        = 4.06      # [eV] electron affinity
       Bgn2Chi  = 0.5       # [1]
       Eg0  = 0.73      # [eV]
       alpha       = 0 # [eV K^-1]
       beta        = 6.3600e+02  # [K]
       Tpar        = 0.0000e+00  # [K]
}
Band2BandTunneling{
    Apath1 = 1.5700e+20     # [1/cm^3/sec]
```

```
    Bpath1  = 5.300e+06       # [V/cm]
    Dpath1 = 0.0000e+00     # [eV]
    Ppath1  = 0.0             # [eV]
    Rpath1  = 0.0000e+00     # [1]
    Apath2 = 1.9500e+16     # [1/cm^3/sec]
    Bpath2  = 1.6700e+07     # [V/cm]
    Dpath2 = 0.0840e+00     # [eV]
    Ppath2  = 0.0060e+00     # [eV]
    Rpath2  = 0.0000e+00     # [1]
    MaxTunnelLength          = 1e-05 # [cm]
}
MultiValley{
    eValley"Gamma"(m0 = 0.0409 energy = 0 degeneracy = 1 alpha0 = 1.35 xid = -7.5) #[1,eV,1,eV^-1]
    eValley"Cond2"(m0 = 0.23 energy = 0.084 degeneracy = 4 alpha0 = 0.73)
    hValley"LH"(m0 = 0.0487 energy = 0 degeneracy = 1 alpha0 = 0.0)
    hValley"HH"(m0 = 0.40 energy = 0 degeneracy = 1 alpha0 = 0.0)
    hValley"SO"(m0 = 0.164 energy = -0.8 degeneracy = 1 alpha0 = 0.0)
}
Epsilon{epsilon = 15.7}  # [1]
}
```

D.3.6 InAs

```
Material="InAs"{
Bandgap{
  * Eg = Eg0 + dEg0 + alpha Tpar^2 / (beta + Tpar) - alpha T^2 / (beta + T)
  * Parameter 'Tpar' specifies the lattice temp at which parameters are defined
            Chi0 = 4.9      # [eV] electron affinity
            Bgn2Chi = 0.5    # [1]
            Eg0 = 0.354   # [eV]
            alpha     = 0    # [eV K^-1]
            beta      = 6.3600e+02    # [K]
            Tpar      = 0.0000e+00    # [K]
}
Band2BandTunneling{
        Apath1 = 1.6500e+20     # [1/cm^3/sec]
        Bpath1  = 1.300e+06       # [V/cm]
        Dpath1 = 0.0000e+00     # [eV]
        Ppath1  = 0.0             # [eV]
        Rpath1  = 0.0000e+00     # [1]
        MaxTunnelLength          = 1e-05 # [cm]
}
MultiValley{
        eValley(0.023, 0.023, 0.023, 0.0000e+00, 1, 2.97)   #[1,1,1,eV,1,eV^-1]
        eValley(0.29, 0.29, 0.29, 0.73, 4, 0.47)
        hValley(0.026, 0.026, 0.026, 0.0000e+00, 1, 0.0)
        hValley(0.41, 0.41, 0.41, 0.0000e+00, 1, 0.0)
        hValley(0.16, 0.16, 0.16, -0.41, 1, 0.0)
}
Epsilon{epsilon = 15.15}  # [1]
}
```

Bibliography

[1] W. G. Vandenberghe, B. Sorée, W. Magnus, G. Groeseneken, and M. V. Fischetti, "Impact of field-induced quantum confinement in tunneling field-effect devices," *Applied Physics Letters*, vol. 98, no. 14, p. 143503, 2011.

[2] A. Vandooren, D. Leonelli, R. Rooyackers, K. Arstila, G. Groeseneken, and C. Huyghebaert, "Impact of process and geometrical parameters on the electrical characteristics of vertical nanowire silicon n-TFETs," *Solid-State Electronics*, vol. 72, pp. 82–87, 2012.

[3] D. Leonelli, A. Vandooren, R. Rooyackers, S. De Gendt, M. Heyns, and G. Groeseneken, "Drive current enhancement in p-tunnel FETs by optimization of the process conditions," *Solid-State Electronics*, vol. 65, pp. 28–32, 2011.

[4] D. Leonelli, A. Vandooren, R. Rooyackers, A. S. Verhulst, S. De Gendt, M. M. Heyns, and G. Groeseneken, "Silicide engineering to boost Si tunnel transistor drive current," *Japanese Journal of Applied Physics*, vol. 50, no. 4S, p. 04DC05, 2011.

[5] W. G. Vandenberghe, A. S. Verhulst, K.-H. Kao, K. De Meyer, B. Sorée, W. Magnus, and G. Groeseneken, "A model determining optimal doping concentration and material's band gap of tunnel field-effect transistors," *Applied Physics Letters*, vol. 100, no. 19, p. 193509, 2012.

[6] D. Mohata, S. Mookerjea, A. Agrawal, Y. Li, T. Mayer, V. Narayanan, A. Liu, D. Loubychev, J. Fastenau, and S. Datta *Applied Physics Express*, Feb.

[7] S. Mookerjea, D. Mohata, T. Mayer, V. Narayanan, and S. Datta, "Temperature-dependent I-V characteristics of a vertical $In_{0.53}$ $Ga_{0.47}As$ tunnel FET," *IEEE Electron Device Letters*, vol. 31, pp. 564–566, June 2010.

[8] H. Zhao, Y. Chen, Y. Wang, F. Zhou, F. Xue, and J. Lee, "InGaAs tunneling field-effect-transistors with atomic-layer-deposited gate oxides," *IEEE Transactions on Electron Devices*, vol. 58, pp. 2990 –2995, Sept. 2011.

[9] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, W. K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. W. Then, and R. Chau, "Fabrication, characterization, and physics of III–V heterojunction tunneling field-effect transistors (H-TFET) for steep sub-threshold swing," *2011 IEEE International Electron Devices Meeting (IEDM)*, vol. 3, pp. 33.6.1–33.6.4, Dec. 2011.

[10] O. Nayfeh, C. Chleirigh, J. Hennessy, L. Gomez, J. Hoyt, and D. Antoniadis, "Design of tunneling field-effect transistors using strained-silicon/strained-germanium type-II staggered heterojunctions," *Electron Device Letters, IEEE*, vol. 29, pp. 1074 –1077, Sept. 2008.

[11] C. Hu, P. Patel, A. Bowonder, K. Jeon, S. H. Kim, W. Y. Loh, C. Y. Kang, J. Oh, P. Majhi, A. Javey, *et al.*, "Prospect of tunneling green transistor for 0.1 V CMOS," *2010 IEEE International Electron Devices Meeting (IEDM)*, pp. 16.1.1–16.1.4, 2010.

[12] W. Vandenberghe, A. Verhulst, G. Groeseneken, B. Soree, and W. Magnus, "Analytical model for point and line tunneling in a tunnel field-effect transistor," in *2008 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 137 –140, Sept. 2008.

[13] D. Verreck, A. Verhulst, K.-H. Kao, W. Vandenberghe, K. De Meyer, and G. Groeseneken, "Quantum mechanical performance predictions of p-n-i-n versus pocketed line tunnel field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 60, no. 7, pp. 2128–2134, 2013.

[14] K.-H. Kao, A. S. Verhulst, W. G. Vandenberghe, B. Soree, W. Magnus, D. Leonelli, G. Groeseneken, and K. De Meyer, "Optimization of gate-on-source-only tunnel FETs with counter-doped pockets," *IEEE Transactions on Electron Devices*, vol. 59, no. 8, pp. 2070–2077, 2012.

[15] P. Solomon, I. Lauer, A. Majumdar, J. Teherani, M. Luisier, J. Cai, and S. Koester, "Effect of uniaxial strain on the drain current of a heterojunction tunneling field-effect transistor," *IEEE Electron Device Letters*, vol. 32, pp. 464–466, Apr. 2011.

[16] P.-F. Guo, L.-T. Yang, Y. Yang, L. Fan, G.-Q. Han, G. Samudra, and Y.-C. Yeo, "Tunneling field-effect transistor: Effect of strain and temperature

on tunneling current," *IEEE Electron Device Letters*, vol. 30, pp. 981–983, Sept. 2009.

[17] F. Conzatti, M. G. Pala, D. Esseni, E. Bano, and L. Selmi, "Strain-induced performance improvements in InAs nanowire tunnel FETs," *IEEE Transactions on Electron Devices*, vol. 59, pp. 2085–2092, Aug. 2012.

[18] M. Van de Put, "Band-to-band tunneling in III-V semiconductor heterostructures," in *2013 IEEE EuroCon*, pp. 2133 – 2139, 2013.

[19] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the sp3d5s* tight-binding formalism: From boundary conditions to strain calculations," *Physical Review B*, vol. 74, no. 20, p. 205323, 2006.

[20] E. Mollick, "Establishing Moore's Law," *IEEE Annals of the History of Computing*, vol. 28, pp. 62–75, July 2006.

[21] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, pp. 82–85, Jan. 1998.

[22] R. Aitken, V. Chandra, J. Myers, B. Sandhu, L. Shifren, and G. Yeric, "Device and technology implications of the internet of things," in *2014 Symposium on VLSI Technology (VLSI-Technology)*, pp. 1–4, 2014.

[23] W. Dehaene and A. Verhulst, "New devices for internet of things: A circuit level perspective," in *2015 IEEE International Electron Devices Meeting (IEDM)*, pp. 25.5.1–25.5.4, Dec. 2015.

[24] A. M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches.," *Nature*, vol. 479, pp. 329–37, Nov. 2011.

[25] A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proceedings of the IEEE*, vol. 98, pp. 2095–2110, Dec. 2010.

[26] H. Lu and A. Seabaugh, "Tunnel field-effect transistors: State-of-the-art," *IEEE Journal of the Electron Devices Society*, vol. 2, pp. 44–49, July 2014.

[27] L. Zhang and M. Chan, *Tunneling Field Effect Transistor Technology*, ch. 2. Springer, 2016.

[28] N. S. Kim, T. Austin, D. Baauw, T. Mudge, K. Flautner, J. S. Hu, M. J. Irwin, M. Kandemir, and V. Narayanan, "Leakage current: Moore's law meets static power," *Computer*, vol. 36, pp. 68–75, Dec. 2003.

[29] R. Dennard, V. Rideout, E. Bassous, and A. LeBlanc, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, pp. 256–268, Oct. 1974.

[30] J. R. Brews, "Subthreshold behavior of uniformly and nonuniformly doped long-channel MOSFET," *IEEE Transactions on Electron Devices*, vol. 26, pp. 1282–1291, Sept. 1979.

[31] S. Salahuddin, , and S. Datta, "Use of negative capacitance to provide voltage amplification for low power nanoscale devices," *Nano Letters*, vol. 8, no. 2, pp. 405–410, 2008.

[32] K. Gopalakrishnan, P. Griffin, and J. Plummer, "I-MOS: a novel semiconductor device with a subthreshold slope lower than kT/q," in *2002 IEEE International Electron Devices Meeting (IEDM)*, pp. 289 – 292, 2002.

[33] V. Pott, H. Kam, R. Nathanael, J. Jeon, E. Alon, and T. J. K. Liu, "Mechanical computing redux: Relays for integrated circuit applications," *Proceedings of the IEEE*, vol. 98, pp. 2076–2094, Dec. 2010.

[34] J. J. Quinn, G. Kawamoto, and B. D. McCombe, "Subband spectroscopy by surface channel tunneling," *Surface Science*, vol. 73, pp. 190–196, May 1978.

[35] S. Banerjee, W. Richardson, J. Coleman, and A. Chatterjee, "A new three-terminal tunnel device," *IEEE Electron Device Letters*, vol. 8, pp. 347–349, Aug. 1987.

[36] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Physical Review Letters*, vol. 93, p. 196805, Nov. 2004.

[37] K. K. Bhuwalka, J. Schulze, and I. Eisele, "Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering," *IEEE Transactions on Electron Devices*, vol. 52, pp. 909–917, May 2005.

[38] B. Rajamohanan, D. Mohata, A. Ali, and S. Datta, "Insight into the output characteristics of III-V tunneling field effect transistors," *Applied Physics Letters*, vol. 102, no. 9, p. 092105, 2013.

[39] A. S. Verhulst, D. Leonelli, R. Rooyackers, and G. Groeseneken, "Drain voltage dependent analytical model of tunnel field-effect transistors," *Journal of Applied Physics*, vol. 110, no. 2, p. 024510, 2011.

[40] L. D. Michielis, L. Lattanzio, and A. M. Ionescu, "Understanding the superlinear onset of tunnel-FET output characteristic," *IEEE Electron Device Letters*, vol. 33, pp. 1523–1525, Nov. 2012.

[41] Q. Smets, D. Verreck, A. S. Verhulst, R. Rooyackers, C. Merckling, M. Van De Put, E. Simoen, W. Vandervorst, N. Collaert, V. Y. Thean, B. Sorée, G. Groeseneken, and M. M. Heyns, "InGaAs tunnel diodes for the calibration of semi-classical and quantum mechanical band-to-band tunneling models," *Journal of Applied Physics*, vol. 115, p. 184503, May 2014.

[42] A. S. Verhulst, D. Verreck, M. A. Pourghaderi, M. Van de Put, B. Sorée, G. Groeseneken, N. Collaert, and A. V.-Y. Thean, "Can p-channel tunnel field-effect transistors perform as good as n-channel?," *Applied Physics Letters*, vol. 105, no. 4, p. 043103, 2014.

[43] K.-H. Kao, A. Verhulst, W. Vandenberghe, B. Soree, G. Groeseneken, and K. de Meyer, "Direct and indirect band-to-band tunneling in germanium-based TFETs," *IEEE Transactions on Electron Devices*, vol. 59, pp. 292–301, Feb. 2012.

[44] D. Griffiths, *Introduction to Quantum Mechanics*. Prentice Hall, 1995.

[45] E. O. Kane, "Theory of tunneling," *Journal of Applied Physics*, vol. 32, no. 1, pp. 83–91, 1961.

[46] J. T. Teherani, "Band-to-band tunneling in silicon diodes and tunnel transistors," *Master thesis, Master of Science in Electrical Engineering and Computer Science*, Massachusetts Institute of Technology, 2010.

[47] E. Kane, "Zener tunneling in semiconductors," *Journal of Physics and Chemistry of Solids*, vol. 12, no. 2, pp. 181–188, 1960.

[48] W. Vandenberghe, B. Sorée, W. Magnus, and G. Groeseneken, "Zener tunneling in semiconductors under nonuniform electric fields," *Journal of Applied Physics*, vol. 107, no. 5, p. 054520, 2010.

[49] K. Ganapathi and S. Salahuddin, "Zener tunneling: Congruence between semi-classical and quantum ballistic formalisms," *Journal of Applied Physics*, vol. 111, no. 12, p. 124506, 2012.

[50] C. Duke, "Theory of metal-barrier-metal tunneling," in *Tunneling phenomena in solids*, pp. 33–35, Springer, 1969.

[51] Synopsys, *Sentaurus Device User Guide*, 2015.06.

[52] M. L. Van de Put, W. G. Vandenberghe, W. Magnus, and B. Sorée, "An envelope function formalism for lattice-matched heterostructures," *Physica B: Condensed Matter*, vol. 470–471, pp. 69 – 75, 2015.

[53] W. Vandenberghe, B. Sorée, W. Magnus, and M. V. Fischetti, "Generalized phonon-assisted Zener tunneling in indirect semiconductors with non-uniform electric fields: A rigorous approach," *Journal of Applied Physics*, vol. 109, no. 12, p. 124503, 2011.

[54] M. G. Burt, "An exact formulation of the envelope function method for the determination of electronic states in semiconductor microstructures," *Semiconductor Science and Technology*, vol. 3, no. 8, p. 739, 1988.

[55] J. C. Slater and G. F. Koster, "Simplified LCAO method for the periodic potential problem," *Physical Review*, vol. 94, pp. 1498–1524, June 1954.

[56] M. P. Anantram, M. S. Lundstrom, and D. E. Nikonov, "Modeling of nanoscale devices," *Proceedings of the IEEE*, vol. 96, pp. 1511–1550, Sept. 2008.

[57] M. Cardona and F. H. Pollak, "Energy-band structure of germanium and silicon: The k.p method," *Physical Review*, vol. 142, pp. 530–543, Feb. 1966.

[58] S. B. Radhia, N. Fraj, I. Saïdi, and K. Boujdaria, "The eight-level k.p model for the conduction and valence bands of InAs, InP, InSb," *Semiconductor Science Technology*, vol. 22, no. 4, p. 427, 2007.

[59] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, "Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters," *Physical Review B*, vol. 57, pp. 6493–6507, Mar. 1998.

[60] C. S. Lent and D. J. Kirkner, "The quantum transmitting boundary method.," *Journal of Applied Physics*, vol. 67, no. 10, p. 6353, 1990.

[61] R. J. LeVeque, *Finite difference methods for ordinary and partial differential equations: steady-state and time-dependent problems*, vol. 98, ch. 1. Siam, 2007.

[62] L. R. Ram-Mohan, *Finite element and boundary element applications in quantum mechanics*, vol. 5. Oxford University Press on Demand, 2002.

[63] L. N. Trefethen, *Spectral methods in MATLAB*, vol. 10. Siam, 2000.

[64] S. Datta, *Quantum Transport: Atom to Transistor*. Cambridge University Press, 2005.

[65] H. Ryu, H. Park, and M. Shin, "Feasibility, accuracy, and performance of contact block reduction method for multi-band simulations of ballistic quantum transport," *Journal of Applied Physics*, vol. 111, no. 6, pp. 1–10, 2012.

[66] P. Long, J. Z. Huang, M. Povolotskyi, G. Klimeck, and M. J. Rodwell, "High-current tunneling FETs with (110) orientation and a channel heterojunction," *IEEE Electron Device Letters*, vol. 37, no. 3, pp. 345–348, 2016.

[67] Q. Zhang, W. Zhao, and A. Seabaugh, "Low-subthreshold-swing tunnel transistors," *IEEE Electron Device Letters*, vol. 27, pp. 297–300, Apr. 2006.

[68] W. G. Vandenberghe, A. S. Verhulst, B. Sorée, W. Magnus, G. Groeseneken, Q. Smets, M. Heyns, and M. V. Fischetti, "Figure of merit for and identification of sub-60 mv/decade devices," *Applied Physics Letters*, vol. 102, no. 1, p. 013510, 2013.

[69] K. Tomioka, M. Yoshimura, and T. Fukui, "Steep-slope tunnel field-effect transistors using III–V nanowire/Si heterojunction," *2012 Symposium on VLSI Technology (VLSIT)*, no. 2010, pp. 47–48, 2012.

[70] F. Mayer, C. L. Royer, J. F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si1-xGexOI and GeOI substrates on CMOS compatible tunnel FET performance," in *2008 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–5, Dec. 2008.

[71] Y. Morita, T. Mori, S. Migita, W. Mizubayashi, A. Tanabe, K. Fukuda, T. Matsukawa, K. Endo, S. O'uchi, Y. X. Liu, M. Masahara, and H. Ota, "Performance enhancement of tunnel field-effect transistors by synthetic electric field effect," *IEEE Electron Device Letters*, vol. 35, pp. 792–794, July 2014.

[72] D. Sarkar, X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, P. M. Ajayan, and K. Banerjee, "A subthermionic tunnel field-effect transistor with an atomically thin channel," *Nature*, vol. 526, pp. 91–95, Jan. 2015.

[73] Q. Huang, R. Huang, C. Wu, H. Zhu, C. Chen, J. Wang, L. Guo, R. Wang, L. Ye, and Y. Wang, "Comprehensive performance re-assessment of TFETs with a novel design by gate and source engineering from device/circuit perspective," in *2014 IEEE International Electron Devices Meeting (IEDM)*, pp. 13.3.1–13.3.4, Dec. 2014.

[74] M. Kim, Y. Wakabayashi, R. Nakane, M. Yokoyama, M. Takenaka, and S. Takagi, "High Ion/Ioff Ge-source ultrathin body strained-SOI tunnel FETs," in *2014 IEEE International Electron Devices Meeting (IEDM)*, pp. 13.2.1–13.2.4, Dec. 2014.

[75] M. G. Pala and D. Esseni, "Interface traps in InAs nanowire tunnel-FETs and MOSFETs - Part I: Model description and single trap analysis in tunnel-FETs," *IEEE Transactions on Electron Devices*, vol. 60, pp. 2795–2801, Sept. 2013.

[76] U. E. Avci, B. Chu-Kung, A. Agrawal, G. Dewey, V. Le, R. Rios, D. H. Morris, S. Hasan, R. Kotlyar, J. Kavalieros, and I. A. Young, "Study of TFET non-ideality effects for determination of geometry and defect density requirements for sub-60mV/dec Ge TFET," in *2015 IEEE International Electron Devices Meeting (IEDM)*, pp. 34.5.1–34.5.4, Dec. 2015.

[77] A. S. Verhulst, D. Verreck, Q. Smets, K. H. Kao, M. V. de Put, R. Rooyackers, B. Sorée, A. Vandooren, K. D. Meyer, G. Groeseneken, M. M. Heyns, A. Mocuta, N. Collaert, and A. V. Y. Thean, "Perspective of tunnel-FET for future low-power technology nodes," in *2014 IEEE International Electron Devices Meeting (IEDM)*, pp. 30.2.1–30.2.4, Dec. 2014.

[78] A. S. Verhulst, W. G. Vandenberghe, K. Maex, and G. Groeseneken, "Tunnel field-effect transistor without gate-drain overlap," *Applied Physics Letters*, vol. 91, no. 5, p. 053102, 2007.

[79] J. V. Morgan and E. O. Kane, "Observation of direct tunneling in germanium," *Physical Review Letters*, vol. 3, pp. 466–468, Nov. 1959.

[80] J. A. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors," *Nature*, vol. 479, no. 7373, pp. 317–323, 2011.

[81] M. Levinshtein, S. Rumyantsev, and M. Schur, *Handbook Series of Semiconductor Parameters*, vol. 1,2. 1999.

[82] S. Koswatta, "On the possibility of obtaining MOSFET-like performance and sub-60-mV/dec swing in 1-D broken-gap tunnel transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 12, pp. 3222–3230, 2010.

[83] J. Knoch and J. Appenzeller, "Modeling of high-performance p-type III–V heterojunction tunnel FETs," *IEEE Electron Device Letters*, vol. 31, no. 4, pp. 305–307, 2010.

[84] P. Yu and M. Cardona, *Fundamentals of Semiconductors*. No. 11-13, Springer, 4th ed., 2010.

[85] H. Ilatikhameneh, Y. Tan, B. Novakovic, G. Klimeck, R. Rahman, and J. Appenzeller, "Tunnel field-effect transistors in 2-D transition metal dichalcogenide materials," *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, vol. 1, pp. 12–18, Dec. 2015.

[86] Q. Zhang, G. Iannaccone, and G. Fiori, "Two-dimensional tunnel transistors based on Bi2Se3 thin film," *IEEE Electron Device Letters*, vol. 35, pp. 129–131, Jan. 2014.

[87] R. K. Ghosh and S. Mahapatra, "Monolayer transition metal dichalcogenide channel-based tunnel transistor," *IEEE Journal of the Electron Devices Society*, vol. 1, pp. 175–180, Oct. 2013.

[88] M. Li, D. Esseni, J. Nahas, D. Jena, and H. Xing, "Two-dimensional heterojunction interlayer tunneling field effect transistors (Thin-TFETs)," *IEEE Journal of the Electron Devices Society*, vol. 3, pp. 200–207, May 2015.

[89] D. Jena, "Tunneling transistors based on graphene and 2-D crystals," *Proceedings of the IEEE*, vol. 101, pp. 1585–1602, July 2013.

[90] Y. Zhang, T.-T. Tang, C. Girit, Z. Hao, M. C. Martin, A. Zettl, M. F. Crommie, Y. R. Shen, and F. Wang, "Direct observation of a widely tunable bandgap in bilayer graphene.," *Nature*, vol. 459, pp. 820–3, June 2009.

[91] F. Xia, V. Perebeinos, Y.-m. Lin, Y. Wu, and P. Avouris, "The origins and limits of metal-graphene junction resistance," *Nature nanotechnology*, vol. 6, no. 3, pp. 179–184, 2011.

[92] V. Nagavarapu, R. Jhaveri, and J. Woo, "The tunnel source (PNPN) n-MOSFET: A novel high performance transistor," *IEEE Transactions on Electron Devices*, vol. 55, pp. 1013–1019, Apr. 2008.

[93] A. Tura, Z. Zhang, P. Liu, Y.-H. Xie, and J. Woo, "Vertical silicon p-n-p-n tunnel nMOSFET with MBE-grown tunneling junction," *IEEE Transactions on Electron Devices*, vol. 58, pp. 1907–1913, July 2011.

[94] Q. Huang, R. Huang, Z. Zhan, Y. Qiu, W. Jiang, C. Wu, and Y. Wang, "A novel Si tunnel FET with 36mV/dec subthreshold slope based on junction depleted-modulation through striped gate configuration," in *2012 IEEE International Electron Devices Meeting (IEDM)*, pp. 8.5.1–8.5.4, Dec. 2012.

[95] K.-H. Kao, A. Verhulst, W. Vandenberghe, and K. De Meyer, "Counterdoped pocket thickness optimization of gate-on-source-only tunnel

FETs," *IEEE Transactions on Electron Devices*, vol. 60, pp. 6–12, Jan. 2013.

[96] K. Boucart and A. Ionescu, "Double-gate tunnel FET with high-k gate dielectric," *IEEE Transactions on Electron Devices*, vol. 54, pp. 1725–1733, July 2007.

[97] R. Rooyackers, A. Vandooren, A. S. Verhulst, A. M. Walke, E. Simoen, K. Devriendt, S. Lo-Corotondo, M. Demand, G. Bryce, R. Loo, A. Hikavyy, T. Vandeweyer, C. Huyghebaert, N. Collaert, and A. V. Y. Thean, "Ge-source vertical tunnel FETs using a novel replacement-source integration scheme," *IEEE Transactions on Electron Devices*, vol. 61, no. 12, pp. 4032–4039, 2014.

[98] A. S. Verhulst, B. Sorée, D. Leonelli, W. G. Vandenberghe, and G. Groeseneken, "Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor," *Journal of Applied Physics*, vol. 107, no. 2, p. 024518, 2010.

[99] M. Luisier and G. Klimeck, "Atomistic full-band design study of InAs band-to-band tunneling field-effect transistors," *IEEE Electron Device Letters*, vol. 30, no. 6, pp. 602–604, 2009.

[100] K. Ganapathi, Y. Yoon, and S. Salahuddin, "Analysis of InAs vertical and lateral band-to-band tunneling transistors: Leveraging vertical tunneling for improved performance," *Applied Physics Letters*, vol. 97, no. 3, p. 033504, 2010.

[101] S. Saurabh and M. Kumar, "Novel attributes of a dual material gate nanoscale tunnel field-effect transistor," *IEEE Transactions on Electron Devices*, vol. 58, pp. 404–410, Feb. 2011.

[102] W. Y. Choi and W. Lee, "Hetero-gate-dielectric tunneling field-effect transistors," *IEEE Transactions on Electron Devices*, vol. 57, pp. 2317–2319, Sept. 2010.

[103] Y. Sun, S. E. Thompson, and T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 101, no. 10, p. 104503, 2007.

[104] M. Chu, Y. Sun, U. Aghoram, and S. E. Thompson, "Strain: A solution for higher carrier mobility in nanoscale MOSFETs," *Annual Review of Materials Research*, vol. 39, pp. 203–229, 2009.

[105] J. Bardeen and W. Shockley, "Deformation potentials and mobilities in non-polar crystals," *Physical Review*, vol. 80, pp. 72–80, Oct. 1950.

[106] G. L. Bir, G. E. Pikus, P. Shelnitz, and D. Louvish, *Symmetry and strain-induced effects in semiconductors*, vol. 624. Wiley New York, 1974.

[107] W. A. Harrison, *Electronic structure and the properties of solids: the physics of the chemical bond*, ch. 7,8. Courier Corporation, 2012.

[108] K.-H. Kao, A. S. Verhulst, M. Van de Put, W. G. Vandenberghe, B. Sorée, W. Magnus, and K. De Meyer, "Tensile strained Ge tunnel field-effect transistors: k.p material modeling and numerical device simulation," *Journal of Applied Physics*, vol. 115, no. 4, p. 044505, 2014.

[109] D. Verreck, A. S. Verhulst, M. L. V. de Put, B. Sorée, N. Collaert, A. Mocuta, A. Thean, and G. Groeseneken, "Uniform strain in heterostructure tunnel field-effect transistors," *IEEE Electron Device Letters*, vol. 37, pp. 337–340, Mar. 2016.

[110] S. Wirths, A. T. Tiedemann, Z. Ikonic, P. Harrison, B. Holländer, T. Stoica, G. Mussler, M. Myronov, J. M. Hartmann, D. Grützmacher, D. Buca, and S. Mantl, "Band engineering and growth of tensile strained Ge/(Si) GeSn heterostructures for tunnel field effect transistors," *Applied Physics Letters*, vol. 102, no. 19, p. 192103, 2013.

[111] M. H. Ben-Jamaa, K. Mohanram, and G. D. Micheli, "An efficient gate library for ambipolar CNTFET logic," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 30, pp. 242–255, Feb. 2011.

[112] X. Yang and K. Mohanram, "Ambipolar electronics," *Rice University TREE1002*, 2010. Available at `https://scholarship.rice.edu/bitstream/handle/1911/27467/ambipolarTREE1002.pdf`.

[113] U. Avci, D. Morris, and I. Young, "Tunnel field-effect transistors: Prospects and challenges," *IEEE Journal of the Electron Devices Society*, vol. 3, pp. 88–95, May 2015.

[114] L. Esaki, "New phenomenon in narrow germanium p-n junctions," *Physical review*, vol. 109, no. 2, p. 603, 1958.

[115] D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S. J. Koester, D. Sylvester, and D. Blaauw, "Low power circuit design based on heterojunction tunneling transistors (HETTs)," in *Proceedings of the 2009 ACM/IEEE international symposium on low power electronics and design*, pp. 219–224, ACM, 2009.

[116] V. Saripalli, S. Datta, V. Narayanan, and J. P. Kulkarni, "Variation-tolerant ultra low-power heterojunction tunnel FET SRAM design," in *Proceedings of the 2011 IEEE/ACM International Symposium on Nanoscale Architectures*, pp. 45–52, IEEE Computer Society, 2011.

[117] D. Morris, U. Avci, R. Rios, and I. Young, "Design of low voltage tunneling-FET logic circuits considering asymmetric conduction characteristics," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 4, pp. 380–388, Dec. 2014.

[118] S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, "On enhanced Miller capacitance effect in interband tunnel transistors," *IEEE Electron Device Letters*, vol. 30, pp. 1102–1104, Oct. 2009.

[119] J. Zhuge, A. S. Verhulst, W. G. Vandenberghe, W. Dehaene, R. Huang, Y. Wang, and G. Groeseneken, "Digital-circuit analysis of short-gate tunnel FETs for low-voltage applications," *Semiconductor Science and Technology*, vol. 26, no. 8, p. 085001, 2011.

[120] N. Dagtekin and A. M. Ionescu, "Impact of super-linear onset, off-region due to uni-directional conductance and dominant Cgd on performance of TFET-based circuits," *IEEE Journal of the Electron Devices Society*, vol. 3, pp. 233–239, May 2015.

[121] A. C. Ford, C. W. Yeung, S. Chuang, H. S. Kim, E. Plis, S. Krishna, C. Hu, and A. Javey, "Ultrathin body InAs tunneling field-effect transistors on Si substrates," *Applied Physics Letters*, vol. 98, no. 11, p. 113105, 2011.

[122] R. Gandhi, Z. Chen, N. Singh, K. Banerjee, and S. Lee, "CMOS-compatible vertical-silicon-nanowire gate-all-around p-type tunneling FETs with-mV/decade subthreshold swing," *IEEE Electron Device Letters*, vol. 32, no. 11, pp. 1504–1506, 2011.

[123] R. Gandhi, Z. Chen, N. Singh, K. Banerjee, and S. Lee, "Vertical Si-nanowire n-type tunneling FETs with low subthreshold swing ($\leq$ 50 mV/decade) at room temperature," *IEEE Electron Device Letters*, vol. 32, no. 4, pp. 437–439, 2011.

[124] R. Li, Y. Lu, G. Zhou, Q. Liu, S. D. Chae, T. Vasen, W. S. Hwang, Q. Zhang, P. Fay, T. Kosel, M. Wistey, H. Xing, and A. Seabaugh, "AlGaSb/InAs tunnel field-effect transistor with on-current of 78 at 0.5 V," *IEEE Electron Device Letters*, vol. 33, no. 3, pp. 363–365, 2012.

[125] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-gate strained-Ge heterostructure tunneling FET (TFET) with record high drive currents and $<$ 60mV/dec subthreshold slope," in *2008 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–3, Dec. 2008.

[126] Y. Lu, S. Bangsaruntip, X. Wang, L. Zhang, Y. Nishi, and H. Dai, "DNA functionalization of carbon nanotubes for ultrathin atomic layer deposition of high-k dielectrics for nanotube transistors with 60 mV/decade switching,"

Journal of the American Chemical Society, vol. 128, no. 11, pp. 3518–3519, 2006.

[127] S. H. Kim, H. Kam, C. Hu, and T. J. K. Liu, "Germanium-source tunnel field effect transistors with record high ION/IOFF," in *2009 Symposium on VLSI Technology (VLSIT)*, pp. 178–179, June 2009.

[128] S. Mookerjea, D. Mohata, R. Krishnan, J. Singh, A. Vallett, A. Ali, T. Mayer, V. Narayanan, D. Schlom, A. Liu, and S. Datta, "Experimental demonstration of 100nm channel length In0.53Ga0.47As-based vertical inter-band tunnel field effect transistors (TFETs) for ultra low-power logic and SRAM applications," in *2009 IEEE International Electron Devices Meeting (IEDM)*, pp. 1–3, Dec. 2009.

[129] G. Zhou, R. Li, T. Vasen, M. Qi, S. Chae, Y. Lu, Q. Zhang, H. Zhu, J. M. Kuo, T. Kosel, M. Wistey, P. Fay, A. Seabaugh, and H. Xing, "Novel gate-recessed vertical InAs/GaSb TFETs with record high ION of 180 uA/um at VDS = 0.5 V," in *2012 IEEE International Electron Devices Meeting (IEDM)*, pp. 32.6.1–32.6.4, Dec. 2012.

[130] D. Leonelli, A. Vandooren, R. Rooyackers, A. S. Verhulst, S. De Gendt, M. M. Heyns, and G. Groeseneken, "Performance enhancement in multi gate tunneling field effect transistors by scaling the fin-width," *Japanese Journal of Applied Physics*, vol. 49, no. 4S, p. 04DC10, 2010.

[131] A. Villalon, C. L. Royer, M. Cassé, D. Cooper, B. Prévitali, C. Tabone, J. M. Hartmann, P. Perreau, P. Rivallin, J. F. Damlencourt, F. Allain, F. Andrieu, O. Weber, O. Faynot, and T. Poiroux, "Strained tunnel FETs with record ION: first demonstration of ETSOI TFETs with SiGe channel and RSD," in *2012 Symposium on VLSI technology (VLSIT)*, pp. 49–50, IEEE, 2012.

[132] K. Jeon, W. Y. Loh, P. Patel, C. Y. Kang, J. Oh, A. Bowonder, C. Park, C. S. Park, C. Smith, P. Majhi, H. H. Tseng, R. Jammy, T. J. K. Liu, and C. Hu, "Si tunnel transistors with a novel silicided source and 46mV/dec swing," in *2010 Symposium on VLSI Technology (VLSIT)*, pp. 121–122, June 2010.

[133] A. Alian, J. Franco, A. Vandooren, Y. Mols, A. Verhulst, S. E. Kazzi, R. Rooyackers, D. Verreck, Q. Smets, A. Mocuta, N. Collaert, D. Lin, and A. Thean, "Record performance InGaAs homo-junction TFET with superior SS reliability over MOSFET," in *2015 IEEE International Electron Devices Meeting (IEDM)*, pp. 31.7.1–31.7.4, Dec. 2015.

[134] B. Ganjipour, J. Wallentin, M. T. Borgstrom, L. Samuelson, and C. Thelander, "Tunnel field-effect transistors based on InP-GaAs

heterostructure nanowires," *ACS nano*, vol. 6, no. 4, pp. 3109–3113, 2012.

[135] A. W. Dey, B. M. Borg, B. Ganjipour, M. Ek, K. A. Dick, E. Lind, P. Nilsson, C. Thelander, and L. E. Wernersson, "High current density InAsSb/GaSb tunnel field effect transistors," in *2012 Annual Device Research Conference (DRC)*, pp. 205–206, June 2012.

[136] P. J. Hasnip, K. Refson, M. I. J. Probert, J. R. Yates, S. J. Clark, and C. J. Pickard, "Density functional theory in the solid state," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 372, no. 2011, 2014.

[137] M. V. Fischetti, B. Fu, S. Narayanan, and J. Kim, *Semiclassical and Quantum Electronic Transport in Nanometer-Scale Structures: Empirical Pseudopotential Band Structure, Monte Carlo Simulations and Pauli Master Equation*, pp. 183–247. New York, NY: Springer New York, 2011.

[138] C. Jacoboni and P. Bordone, "The wigner-function approach to non-equilibrium electron transport," *Reports on Progress in Physics*, vol. 67, no. 7, p. 1033, 2004.

[139] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III-V compound semiconductors and their alloys," *Journal of Applied Physics*, vol. 89, no. 11, p. 5815, 2001.

[140] Y. X. Liu, D. Z. Y. Ting, and T. C. McGill, "Efficient, numerically stable multiband k.p treatment of quantum transport in semiconductor heterostructures," *Physical Review B*, vol. 54, pp. 5675–5683, Aug. 1996.

[141] M. G. Burt, "An exact formulation of the envelope function method for the determination of electronic states in semiconductor microstructures," *Semiconductor Science and Technology*, vol. 2, no. 7, p. 460, 1987.

[142] D. Verreck, M. Van de Put, B. Sorée, A. S. Verhulst, W. Magnus, W. G. Vandenberghe, N. Collaert, A. Thean, and G. Groeseneken, "Quantum mechanical solver for confined heterostructure tunnel field-effect transistors," *Journal of Applied Physics*, vol. 115, no. 5, p. 053706, 2014.

[143] P. Yu and M. Cardona, *Fundamentals of Semiconductors*, ch. 2, pp. 46–48. Springer, 4th ed., 2010.

[144] M. Burt, "Resolution of the out-of-zone solution problem in envelope-function theory," *Superlattices and microstructures*, vol. 23, no. 2, pp. 531–534, 1998.

[145] D. Griffiths, *Introduction to Quantum Mechanics*. Prentice Hall, 1995.

[146] M. Schuurmans and G. 't Hooft, "Simple calculations of confinement states in a quantum well," *Physical Review B*, vol. 31, no. 12, p. 8041, 1985.

[147] S. White and L. Sham, "Electronic properties of flat-band semiconductor heterostructures," *Physical Review Letters*, vol. 47, no. 12, p. 879, 1981.

[148] J. P. Boyd, *Chebyshev and Fourier spectral methods*. Courier Corporation, 2001.

[149] A. J. Jerri, *The Gibbs phenomenon in Fourier analysis, splines and wavelet approximations*, vol. 446, pp. 1–16. Springer Science & Business Media, 2013.

[150] M. Van de Put, "Modeling of quantum electron transport with applications in energy filtering nanostructures," *PhD Thesis, PhD in Science: Physics*, University of Antwerp, 2016.

[151] S. Martucci, "Symmetric convolution and the discrete sine and cosine transforms," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1038–1051, May 1994.

[152] N. Fraj, I. Saïdi, S. Ben Radhia, and K. Boujdaria, "Band structures of AlAs, GaP, and SiGe alloys: A 30 k.p model," *Journal of Applied Physics*, vol. 102, no. 5, p. 053703, 2007.

[153] I. Saïdi, S. Ben Radhia, and K. Boujdaria, "Band parameters of GaAs, InAs, InP, and InSb in the 40-band k.p model," *Journal of Applied Physics*, vol. 107, no. 4, p. 043701, 2010.

[154] D. Griffiths, *Introduction to Quantum Mechanics*. Prentice Hall, 1995.

[155] S. Johnson, "The NLopt nonlinear-optimization package." Available at `http://ab-initio.mit.edu/nlopt`.

[156] N. T. Bliss and J. Kepner, "'pMATLAB parallel MATLAB library'," *International Journal of High Performance Computing Applications*, vol. 21, no. 3, pp. 336–359, 2007.

[157] Synopsys, *Sentaurus Device User Guide*, 2013.12.

[158] E. Baravelli, E. Gnani, R. Grassi, A. Gnudi, S. Reggiani, and G. Baccarani, "Optimization of n- and p-type TFETs integrated on the same InAs/$Al_xGa_{1-x}Sb$ technology platform," *IEEE Transactions on Electron Devices*, vol. 61, pp. 178–185, Jan. 2014.

[159] U. Avci, R. Rios, K. Kuhn, and I. Young, "Comparison of power and performance for the TFET and MOSFET and considerations for P-TFET," in *2011 11th IEEE Conference on Nanotechnology (IEEE-NANO)*, pp. 869–872, Aug. 2011.

[160] D. Yakimets, "Vertical transistors: A slippery path towards the ultimate cmos scaling," *PhD Thesis, PhD in Engineering Science: Electrical Engineering*, KU Leuven, 2016.

[161] D. Verreck, A. S. Verhulst, M. Van de Put, B. Sorée, W. Magnus, A. Mocuta, N. Collaert, A. Thean, and G. Groeseneken, "Full-zone spectral envelope function formalism for the optimization of line and point tunnel field-effect transistors," *Journal of Applied Physics*, vol. 118, no. 13, p. 134502, 2015.

[162] D. Verreck, A. S. Verhulst, B. Sorée, N. Collaert, A. Mocuta, A. Thean, and G. Groeseneken, "Improved source design for p-type tunnel field-effect transistors: Towards truly complementary logic," *Applied Physics Letters*, vol. 105, no. 24, p. 243506, 2014.

[163] P. Yu and M. Cardona, *Fundamentals of Semiconductors*, ch. 2, pp. 71–75. Springer, 4th ed., 2010.

[164] M. Cardona, K. L. Shaklee, and F. H. Pollak, "Electroreflectance at a semiconductor-electrolyte interface," *Physical Review*, vol. 154, no. 3, p. 696, 1967.

[165] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Physical Review B*, vol. 14, no. 2, p. 556, 1976.

[166] P. Yu and M. Cardona, *Fundamentals of Semiconductors*, ch. Appendix B, pp. 601–604. Springer, 4th ed., 2010.

[167] T. B. Bahder, "Eight-band k.p model of strained zinc-blende crystals," *Physical Review B*, vol. 41, no. 17, p. 11992, 1990.

[168] R. Neffati, I. Saidi, and K. Boujdaria, "Full-zone k.p model for the electronic structure of unstrained $GaAs_{1-x}P_x$ and strained $Al_xIn_{1-x}As$ alloys," *Journal of Applied Physics*, vol. 112, no. 5, p. 053716, 2012.

[169] R. Kotlyar, U. E. Avci, S. Cea, R. Rios, T. D. Linton, K. J. Kuhn, and I. A. Young, "Bandgap engineering of group IV materials for complementary n and p tunneling field effect transistors," *Applied Physics Letters*, vol. 102, no. 11, p. 113106, 2013.

[170] L. Knoll, Q.-T. Zhao, A. Nichau, S. Trellenkamp, S. Richter, A. Schafer, D. Esseni, L. Selmi, K. Bourdelle, and S. Mantl, "Inverters with strained Si nanowire complementary tunnel field-effect transistors," *IEEE Electron Device Letters*, vol. 34, pp. 813–815, June 2013.

[171] G. Han, Y. Wang, Y. Liu, H. Wang, M. Liu, C. Zhang, J. Zhang, B. Cheng, and Y. Hao, "Relaxed germanium-tin p-channel tunneling field-effect transistors fabricated on Si: impacts of Sn composition and uniaxial tensile strain," *AIP Advances*, vol. 5, no. 5, p. 057145, 2015.

[172] S. Brocard, M. G. Pala, and D. Esseni, "Design options for hetero-junction tunnel FETs with high on current and steep sub-threshold voltage slope," in *2013 IEEE International Electron Devices Meeting (IEDM)*, pp. 5.4.1–5.4.4, IEEE, Dec. 2013.

[173] R. A. Minamisawa, M. J. Süess, R. Spolenak, J. Faist, C. David, J. Gobrecht, K. K. Bourdelle, and H. Sigg, "Top-down fabricated silicon nanowires under tensile elastic strain up to 4.5%," *Nature Communications*, vol. 3, p. 1096, Oct. 2012.

[174] N. Waldron, C. Merckling, L. Teugels, P. Ong, S. Ibrahim, F. Sebaai, A. Pourghaderi, K. Barla, N. Collaert, and A.-Y. Thean, "InGaAs gate-all-around nanowire devices on 300mm Si substrates," *IEEE Electron Device Letters*, vol. 35, pp. 1097–1099, Nov. 2014.

[175] Synopsys, *Sentaurus Process User Guide*, 2014.12.

[176] M. Levinshtein and S. Rumyantsev, *Handbook Series of Semiconductor Parameters*, vol. 1, ch. 4. 1999.

[177] Y. A. Vul, *Handbook Series of Semiconductor Parameters*, vol. 1, ch. 6. 1999.

[178] M. P. Mikhailova, *Handbook Series of Semiconductor Parameters*, vol. 1, ch. 7. 1999.

[179] M. W. Larsson, J. B. Wagner, M. Wallin, P. Håkansson, L. E. Fröberg, L. Samuelson, and L. R. Wallenberg, "Strain mapping in free-standing heterostructured wurtzite InAs/InP nanowires," *Nanotechnology*, vol. 18, no. 1, p. 15504, 2007.

[180] C. Wu, R. Huang, Q. Huang, J. Wang, and Y. Wang, "Design guideline for complementary heterostructure tunnel FETs with steep slope and improved output behavior," *IEEE Electron Device Letters*, vol. 37, no. 1, pp. 20–23, 2016.

[181] D. Verreck, A. S. Verhulst, B. Sorée, N. Collaert, A. Mocuta, A. Thean, and G. Groeseneken, "Non-uniform strain in lattice-mismatched heterostructure tunnel field-effect transistors," pp. 412–415, Sept.

[182] A. Das and A. C. Melissinos, *Quantum Mechanics: a modern introduction*, pp. 570–571. CRC Press, 1986.

[183] D. A. Miller, *Quantum mechanics for scientists and engineers*, pp. 138–152. Cambridge University Press, 2008.

[184] R. Landauer, "Conductance from transmission: common sense points," *Physica Scripta*, vol. 1992, no. T42, p. 110, 1992.

[185] S. E. Laux, "Analysis of quantum ballistic electron transport in ultrasmall silicon devices including space-charge and geometric effects," *Journal of Applied Physics*, vol. 95, no. 10, p. 5545, 2004.

[186] W. Potz, "Selfconsistent model of transport in quantum well tunneling structures," *Journal of Applied Physics*, vol. 66, no. 6, pp. 2458–2466, 1989.

[187] D. Sels, B. Sorée, and G. Groeseneken, "Quantum ballistic transport in the junctionless nanowire pinch-off field effect transistor," *Journal of Computational Electronics*, vol. 10, no. 1-2, pp. 216–221, 2011.

[188] D. Vasileska, S. Goodnick, and G. Klimeck, *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*, ch. 4.1.7. 2010.

[189] D. Vasileska, S. Goodnick, and G. Klimeck, *Computational Electronics: Semiclassical and Quantum Device Modeling and Simulation*, ch. 4.1.4. 2010.

[190] A. Trellakis, A. Galick, A. Pacelli, and U. Ravaioli, "Iteration scheme for the solution of the two-dimensional Schrödinger-Poisson equations in quantum structures," *Journal of Applied Physics*, vol. 81, no. 12, pp. 7880–7884, 1997.

[191] S. Steiger, M. Povolotskyi, H. H. Park, T. Kubis, and G. Klimeck, "NEMO5: A parallel multiscale nanoelectronics modeling tool," *IEEE Transactions on Nanotechnology*, vol. 10, pp. 1464–1474, Nov. 2011.

[192] Y. He, T. Kubis, M. Povolotskyi, J. Fonseca, and G. Klimeck, "Quantum transport in NEMO5: Algorithm improvements and high performance implementation," in *2014 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 361–364, IEEE, 2014.

[193] J. Charles, P. Sarangapani, R. Golizadeh-Mojarad, R. Andrawis, D. Lemus, X. Guo, D. Mejia, J. E. Fonseca, M. Povolotskyi, T. Kubis, *et al.*, "Incoherent transport in NEMO5: Realistic and efficient scattering on phonons," *Journal of Computational Electronics*, pp. 1–7, 2016.

[194] P. Yu and M. Cardona, *Fundamentals of Semiconductors*, ch. 3, pp. 131–135. Springer, 4th ed., 2010.

[195] C. Jacoboni and L. Reggiani, "The Monte-Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, vol. 55, no. 3, pp. 645–705, 1983.

[196] Y.-M. Niquet, V.-H. Nguyen, F. Triozon, I. Duchemin, O. Nier, and D. Rideau, "Quantum calculations of the carrier mobility: Methodology, Matthiessen's rule, and comparison with semi-classical approaches," *Journal of Applied Physics*, vol. 115, no. 5, p. 054512, 2014.

Curriculum Vitae

Devin Verreck was born in Leuven (Belgium) in 1989. He received the BSc degree in Electrical Engineering from the KU Leuven (Belgium) magna cum laude in 2009 and the MSc degree in Nanoscience and Nanotechnology, option Engineering summa cum laude from the KU Leuven (Belgium) in 2012. Since 2012, he has been pursuing a PhD degree from the KU Leuven, doing his research at imec (Belgium) while being connected to the Department of Electrical Engineering under the supervision of dr. Anne Verhulst, prof. Bart Sorée and prof. Guido Groeseneken. In 2016, he was a visiting researcher with the Network for Computational Nanotechnology at Purdue University in Indiana (USA), under the supervision of prof. Mykhailo Povolotskyi and prof. Gerhard Klimeck.

He was laureate of the Flemish Geo-Olympiad (2007) and the recipient of a PhD stipend from the Institute for the Promotion of Innovation through Science and Technology in Flanders (IWT-Vlaanderen) from 2013 to 2016. His research interests include the quantum mechanical modeling of new nano-electronic device concepts, particularly the tunnel field-effect transistor.

List of publications

Journal papers

First author

1. **D. Verreck**, G. Groeseneken and A.S. Verhulst "The tunnel field-effect transistor", *Wiley Encyclopedia of Electrical and Electronics Engineering*, 1-24, Nov. 2016.

2. **D. Verreck**, A.S. Verhulst, M.L. Van de Put, B. Sorée, A. Mocuta, N. Collaert, A. Thean and G. Groeseneken, "Uniform strain in heterostructure tunnel field-effect transistors", *IEEE Electron Device Letters*, vol. 37, no. 3, pp. 337-340, Mar. 2016.

3. **D. Verreck**, A.S. Verhulst, M.L. Van de Put, B. Sorée, W. Magnus, A. Mocuta, N. Collaert, A. Thean and G. Groeseneken, "Full-zone spectral envelope function formalism for the optimization of line and point tunnel field-effect transistors", *Journal of Applied Physics* 118, 134502, Oct. 2015.

4. **D. Verreck**, A.S. Verhulst, B. Sorée, N. Collaert, A. Mocuta, A. Thean and G. Groeseneken, "Improved source design for p-type tunnel field-effect transistors: Towards truly complementary logic", *Applied Physics Letters*, vol. 105, 243506, Dec. 2014.

5. **D. Verreck**, M.L. Van de Put, B. Sorée, A.S. Verhulst, W. Magnus, W.G. Vandenberghe, N. Collaert, A. Thean and G. Groeseneken, "Quantum mechanical solver for confined heterostructure tunnel field-effect transistors", *Journal of Applied Physics*, vol. 115, no.5, pp.053706-8, Feb. 2014.

6. **D. Verreck**, A.S. Verhulst, K-H. Kao, W.G. Vandenberghe, K. De Meyer, G. Groeseneken, "Quantum Mechanical Performance Predictions of p-n-i-n Versus Pocketed Line Tunnel Field-Effect Transistors", *IEEE Transactions on Electron Devices*, vol. 60, no. 7, pp. 2128-2134, July 2013.

Co-author

1. A. Alian, Y. Mols, C. C. M. Bordallo, **D. Verreck**, A. Verhulst, A. Vandooren, R. Rooyackers, A. Thean, D. Lin, D. Mocuta, N. Collaert, "InGaAs tunnel FET with sub-nanometer EOT and sub-60 mV/dec sub-threshold swing at room temperature", *Applied Physics Letters*, vol. 109, no. 24, 243502, 2016.

2. M. Mohammed, A.S. Verhulst, **D. Verreck**, R. Degraeve, B. Kaczer, E. Simoen, B. Sorée, M.L. Van de Put, A. Mocuta, N. Collaert, A. Thean and G. Groeseneken, "Electric-field induced quantum broadening of the characteristic energy level of traps in semiconductors and oxides", *accepted to Journal of Applied Physics*, 2016.

3. Q. Smets, A.S. Verhulst, K. Martens, D. Lin, S. El Kazzi, **D. Verreck**, E. Simoen, N. Collaert, A. Thean, J-P. Raskin and M.M. Heyns, "Impact of field-induced quantum confinement on the onset of tunneling field-effect transistors: Experimental verification", *Applied Physics Letters*, vol. 150, no. 20, 203507, 2014

4. A.S. Verhulst, **D. Verreck**, A. Pourghaderi, M.L. Van de Put, B. Sorée, G. Groeseneken, N. Collaert and A. Thean, "Can p-channel tunnel-field-effect transistors perform as good as n-channel tunnel-FETs?" *Applied Physics Letters*, vol. 105, no. 4, 43103, 2014.

5. Q. Smets, **D. Verreck**, A.S. Verhulst, R. Rooyackers, C. Merckling, M.L. Van de Put, E. Simoen, W. Vandervorst, N. Collaert, A. Thean, B. Sorée, G. Groeseneken and M.M. Heyns, "InGaAs tunnel diodes for the calibration of semi-classical and quantum mechanical band-to-band tunneling models", *Journal of Applied Physics*, vol. 115, no. 18, 184503, 1-9, 2014.

6. A. Walke, A.S. Verhulst, A. Vandooren, **D. Verreck**, E. Simoen, V. Rao, G. Groeseneken, N. Collaert, A. Thean, "Part I: Impact of Field Induced Quantum Confinement on Trap Assisted Tunneling in Line TFETs.", *IEEE*

Transactions on Electron Devices, vol. 60, no. 12, 4057-4064, 2013.

Patent applications

1. **D. Verreck** and A.S. Verhulst, LAYERED STRUCTURE OF A P-TFET, *US Patent App.* 20160104769, Apr. 2016.

2. A.S. Verhulst, **D. Verreck** and A. Alian, DRAIN EXTENSION REGION FOR TFET, submitted Dec. 2015.

Conference papers

First author

1. **D. Verreck**, A.S. Verhulst, B. Sorée, N. Collaert, A. Mocuta, A. Thean and G. Groeseneken, "Non-uniform strain in lattice-mismatched heterostructure tunnel field-effect transistors", *European Solid-State Device Research Conference (ESSDERC)*, Lausanne, Switzerland, Sept. 2016.

2. **D. Verreck**, M.L. Van de Put, A.S. Verhulst, B. Sorée, W. Magnus, A. Dabral, A. Thean and G. Groeseneken, "15-Band spectral envelope function formalism applied to broken gap tunnel field-effect transistors", *International Workshop on Computational Electronics (IWCE)*, West Lafayette IN, USA, Sept. 2015.

3. **D. Verreck**, M.L. Van de Put, B. Sorée, A.S. Verhulst, W. Magnus, W.G. Vandenberghe, Q. Smets, R. Rooyackers, N. Collaert, A. Thean, M.M. Heyns and G. Groeseneken, "Quantum mechanical modeling of tunnel field-effect transistors for low power nano-electronics", *10th International Nanotechnology Conference (INC10)*, Gaithersburg MD, USA, May 2014. **Winner of Best Poster Award.**

4. **D. Verreck**, M.L. Van de Put, B. Sorée, A.S. Verhulst, W. Magnus, W.G. Vandenberghe, N. Collaert, A. Thean and G. Groeseneken, "Quantum mechanical solver for confined heterostructure tunnel field-effect transistors", *APS March Meeting*, vol 59, 1, Denver CO, USA, 2014.

5. **D. Verreck**, A.S. Verhulst and G. Groeseneken, "Split pocket p-n-i-n tunnel field-effect transistors", *14th International Conference on Ultimate Integration On Silicon (ULIS)*, Warwick UK, Mar. 2013.

Co-author

1. P. Long, J.Z. Huang, M. Povolotskyi, **D. Verreck**, J. Charles, T. Kubis, G. Klimeck, M.J.W. Rodwell, B.H. Calhoun, "A tunnel FET Design for high-Current, 120 mV Operation", *IEEE International Electron Devices Meeting (IEDM)*, San Francisco CA, USA, Dec. 2016.

2. A.S. Verhulst, Q. Smets, J. Bizindavyi, **D. Verreck**, S. El Kazzi, A. Alian, J. Franco, Y. Mols, A. Vandooren, R. Rooyackers, D. Lin, A. Mocuta, B. Sorée, G. Groeseneken, N. Collaert and M.M. Heyns, "Perspective on III-V Tunnel-FETs: bridging the gap between ideal device design and experimental realizations through calibration (invited)", *47th IEEE Semiconductor Interface Specialists Conference (SISC)*, San Diego CA, USA, Dec. 2016.

3. A. Vandooren, A. Alian, A.S. Verhulst, J. Franco, R. Rooyackers, Q. Smets, **D. Verreck**, N. Waldron, D. Mocuta and N. Collaert, "Tunnel FETs for low power electronics (invited)", *IEEE SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S)*, San Francisco CA, USA, Oct. 2016.

4. P. Long, J.Z. Huang, M. Povolotskyi, **D. Verreck**, G. Klimeck, M.J.W. Rodwell, "High-current InP-based triple heterojunction tunnel transistors", *28th International Conference on Indium Phosphide and Related Materials (IPRM)*, Toyama, Japan, Jun. 2016.

5. N. Collaert, A. Alian, H. Arimura, G. Boccardi, G. Eneman, J. Franco, T. Ivanov, D. Lin, J. Mitard, S. Ramesh, R. Rooyackers, M. Schaekers, A. Sibaja-Hernandez, S. Sioncke, Q. Smets, A. Vais, A. Vandooren, A. Veloso, A.S. Verhulst, **D. Verreck**, N. Waldron, A. Walke, L. Witters, H. Yu, D. Zhou and A. Thean, "Beyond-Si materials and devices for more Moore and more than Moore applications", *International Conference on IC Design and Technology (ICICDT)*, Ho Chi Minh City, Vietnam, Jun. 2016.

6. A.S. Verhulst, A. Alian, R. Rooyackers, **D. Verreck**, Q. Smets, Y. Mols, J. Franco, K. Salimi, A. Vandooren, M.L. Van de Put, B. Sorée,

G. Groeseneken, M.M. Heyns, A. Mocuta, D. Lin, N. Collaert and A. Thean, "Tunnel field-effect transistors for low-power nano-electronics", *12th International Nanotechnology Conference (INC12)*, Leuven, Belgium, May 2016.

7. A. Alian, A. Vandooren, Y. Mols, A.S. Verhulst, S. El Kazzi, R. Rooyackers, **D. Verreck**, Q. Smets, A. Mocuta, N. Collaert, D. Lin and A. Thean, "Record performance InGaAs homo-junction TFET with superior SS reliability over MOSFET", *International Electron Devices Meeting (IEDM)*, Washington DC, USA, Dec. 2015.

8. A. Vandooren, R. Rooyackers, A. Alian, A.S. Verhulst, **D. Verreck**, Q. Smets, A. Mocuta, N. Collaert and A. Thean, "Trends and Challenges in Tunnel Field Effect Transistors", *8th International NATO Advanced Workshop*, Kiev, Ukraine, 2015.

9. A.S. Verhulst, Q. Smets, **D. Verreck**, M.L. Van de Put, R. Rooyackers, A. Vandooren, S. El Kazzi, A. Alian, B. Sorée, A. Mocuta, G. Groeseneken, M.M. Heyns, N. Collaert and A. Thean, "The future of tunnel field-effect transistors (invited)", *Nanoelectronics Days*, Juelich, Germany, Apr. 2015.

10. A.S. Verhulst, **D. Verreck**, Q. Smets, K-H. Kao, M.L. Van de Put, R. Rooyackers, B. Sorée, A. Vandooren, K. De Meyer, G. Groeseneken, M.M. Heyns, A. Mocuta, N. Collaert and A. Thean, "Perspective of tunnel-FET for future low-power technology nodes (invited)", *International Electron Devices Meeting (IEDM)*, San Francisco CA, USA, Dec. 2014.

11. Q. Smets, A.S. Verhulst, D. Lin, **D. Verreck**, C. Merckling, S. El Kazzi, K. Martens, J-P. Raskin, A. Thean and M.M. Heyns, "Band-to-band tunneling MOSCAPs for rapid TFET characterization", *72nd Annual Device Research Conference (DRC)*, Santa Barbara CA, USA, 2014.

12. Q. Smets, A.S. Verhulst, R. Rooyackers, C. Merckling, D. Lin, E. Simoen, A. Alian, C. Mirco, A. Pourghaderi, K-H. Kao, **D. Verreck**, N. Collaert, A. Thean and M.M. Heyns, "In0.53Ga0.47As diodes for band-to-band tunneling calibration: design, fabrication and characterization", *Solid State Devices and Materials Conference (SSDM)*, Fukuoka, Japan, 2013.

FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
IMEC/MICAS
Kapeldreef 75
B-3001 Leuven